ESCUELA SUPERIOR DE INGENIERÍA

INGENIERÍA TÉCNICA EN INFORMÁTICA DE GESTIÓN

Generador de casos de prueba aleatorio basado en especificaciones abstractas

Miguel Ángel Pérez Montero

20 de febrero de 2012

ESCUELA SUPERIOR DE INGENIERÍA

INGENIERO TÉCNICO EN INFORMÁTICA DE GESTIÓN

Generador de casos de prueba aleatorio basado en especificaciones abstractas

- Departamento: Lenguajes y Sistemas Informáticos.
- Directores del proyecto: Juan José Dominguéz Jiménez, Antonio García Dominguez.
- Autor del proyecto: Miguel Ángel Pérez Montero.

Cádiz, 20 de febrero de 2012

Fdo: Miguel Ángel Pérez Montero

Licencia

Notación y formato

En esta sección, incluiremos los aspectos relevantes a la notación y el formato a lo largo del documento. Dicha notación es la siguiente:

Si nos vamos a referir a un directorio en particular, usaremos la notación:

/home

Cuando nos refiramos a un programa en concreto, utilizaremos la notación:

emacs.

Cuando nos refiramos a un comando, o función de un lenguaje, usaremos la notación:

`quicksort`.

Si nos vamos a referir a una clase, utilizaremos la notación:

MiClase

Índice general

Indice de figuras

Indice de listados

Capítulo 1

Motivación y contexto

1.1. Introducción

Dentro de la línea de pruebas del software del grupo de investigación UCASE de la Universidad de Cádiz, se trabaja en la mejora de conjuntos de casos de prueba para composiciones de servicios Web escritas en WS-BPEL. Para mejorar el conjunto de casos de prueba, el grupo de investigación propone localizar aquellos posibles fallos que el conjunto de casos de partida no detectaría, y extenderlo para que sí los detecte. Actualmente, el grupo ha obtenido buenos resultados localizando los fallos usando análisis de mutaciones (con sus herramientas *MuBPEL* [1] y *GAmera* [2]), y ha comenzado a trabajar en el área de generación de casos de prueba. Sin embargo, extender un conjunto de casos de prueba para una composición WS-BPEL es difícil, ya que hay que tener en cuenta numerosas tecnologías.

En un Proyecto Fin de Carrera anterior, "Analizador de Servicios Web basados en WSDL 1.1 para pruebas paramétricas", se implementó *ServiceAnalyzer* [3] un analizador de servicios Web, que generaba plantillas parametrizadas para producir mensajes de acuerdo a todas las restricciones impuestas desde WSDL [4], XML Schema [5] y el WS-I Basic Profile 1.1 [6]. Con eso se tenía parte del trabajo realizado, pero ahora faltaba otra cuestión: producir los datos con que se rellenarán esas plantillas.

Como una primera aproximación, de cara a una futura comparación con estrategias

más avanzadas, el grupo decidió realizar una prueba de concepto generando aleatoriamente los casos de prueba. Así nace la idea de este Proyecto Fin de Carrera: *TestGenerator*, una aplicación que es capaz de generar los datos necesarios de forma aleatoria con unas restricciones específicas dadas en un fichero de entrada al programa. Por su simplicidad de análisis, en este PFC se generarán los valores de acuerdo a una distribución aleatoria uniforme.

1.2. Objetivos

La mayoría de los objetivos que debe cumplir *TestGenerator* están motivados por otro programa del grupo llamado *ServiceAnalyzer*. Como ya se ha mencionado anteriormente, *ServiceAnalyzer* es un analizador de servicios Web basados en WSDL 1.1 para pruebas paramétricas. *TestGenerator* deberá dar respaldo a todos los tipos de datos y restricciones que incluye *ServiceAnalyzer*.

TestGenerator debe de ser capaz de dar soporte para los siguientes tipos de datos:

- Enteros
- Reales
- Cadenas
- Fechas
- Horas
- Fechas y horas
- Duraciones
- Listas
- N-Tuplas

Según el tipo de dato, *TestGenerator* deberá dar soporte a ciertas restricciones:

- Enteros: darán soporte a restricciones de valor máximo y mínimo, además de poder tener un número total de dígitos, o bien tomar una lista de valores válidos.

- Reales: tendrán las mismas que los Enteros, pero también el número total de dígitos decimales.

- Cadenas: podrán tomar una lista de valores válidos o bien deberá cumplir cierta expresión regular. También podrán tener una longitud de cadena mínimas y máxima.

- Fechas, Horas, Fechas y horas y Duraciones: tendrán soporte para las restricciones de valor máximo y mínimo.

- Listas y N-Tuplas: estas clases contenedores, deberán tener soporte para especificar qué tipo de elemento contendrán así como la capacidad máxima y mínima.

La aplicación deberá ser capaz de aceptar varios formatos de especificaciones abstractas, y se estructurará de forma que se puedan añadir nuevos formatos en el futuro. Uno de los formatos de entrada serán los catálogos de plantillas que produce *ServiceAnalyzer*, permitiendo generar una serie de valores que produzcan un determinado tipo de mensajes de un servicio Web (entradas, salidas o errores).

Para completar la aplicación, *TestGenerator* también podrá recibir una serie de definiciones de tipos y declaraciones de variables en un lenguaje de dominio específico [7] textual más cómodo que el formato XML generado por *ServiceAnalyzer*. Los tipos y restricciones disponibles serán un superconjunto de los de *ServiceAnalyzer*.

Por otro lado, los datos generados deberán presentarse en un formato que sea utilizable por las herramientas del grupo de investigación. Para ello, deberán ser compatibles con *BPELUnit*, el marco de pruebas unitarias para WS-BPEL que se emplea en el grupo. Como mínimo, se deberán generar ficheros en estos dos formatos:

- Valores separados por comas (Comma Separated Values o CSV): un fichero de texto plano en que cada fila de datos se representa como una línea, dividida en cam-

pos mediante comas. Este formato es fácil de entender y utilizar, pero no admite describir valores con estructuras complejas (listas anidadas dentro de listas).

- Fragmentos de plantillas Apache Velocity, en la que se asignan una serie de valores a las variables que se emplearán dentro de BPELUnit. La sintaxis de Velocity permite describir valores con estructuras de cualquier complejidad, pudiendo incluir árboles completos de valores de tipos heterogéneos.

1.3. Alcance

Este proyecto será capaz de dar respaldo al sistema de tipos que implementa *ServiceAnalyzer*. Este dominio cubre las necesidades del grupo UCASE, ya que fue escogido en base a las composiciones WS-BPEL que se están estudiando actualmente.

En este Proyecto Fin de Carrera, se implementará sólo un generador de datos uniformemente distribuido, dejando de lado otras distribuciones de probabilidad. La distribución aleatoria uniforme es la más comúnmente usada como base de comparación en la literatura de generación de casos de prueba.

La comunicación entre el usuario y la herramienta se hará a través de la línea de órdenes. Por restricciones de tiempo, no se implementará una interfaz gráfica para la herramienta.

1.4. Conceptos básicos

Para poder entender bien cómo encaja este proyecto en el grupo de investigación UCASE es necesario entender algunas de las tecnologías que utiliza dicho grupo, puesto que de forma directa o indirecta afecta a los requisitos de la aplicación así como a su desarrollo.

1.4.1. WS-BPEL

WS-BPEL (Business Process Execution Language) es el lenguaje en el que se centran algunas de las líneas de trabajo del grupo UCASE. WS-BPEL es un lenguaje para la composición de servicios Web. Está basado en XML y sirve para el control centralizado de la invocación de diferentes servicios Web, con cierta lógica de negocio añadida que ayuda a la programación en gran escala.

La estructura de un proceso WS-BPEL se divide en cuatro secciones:

1. Definición de relaciones con los socios externos, que son el cliente que utiliza el proceso de negocio y los WS a los que llama el proceso.

2. Definición de las variables que emplea el proceso.

3. Definición de los distintos tipos de manejadores que puede utilizar el proceso. Pueden definirse manejadores de fallos, que indican las acciones a realizar en caso de producirse un fallo interno o en un WS al que se llama. También se definen los manejadores de eventos, que especifican las acciones a realizar en caso de que el proceso reciba una petición durante su flujo normal de ejecución.

4. Descripción del comportamiento del proceso de negocio; esto se logra a través de las actividades que proporciona el lenguaje.

Todos los elementos definidos anteriormente son globales si se declaran dentro del proceso. Sin embargo, también existe la posibilidad de declararlos de forma local mediante el contenedor scope, que permite dividir el proceso de negocio en diferentes ámbitos. Los principales elementos constructivos de un proceso WS-BPEL son las actividades, que pueden ser de dos tipos: básicas y estructuradas. Las actividades básicas son las que realizan una determinada labor (recepción de un mensaje, manipulación de datos, etc.). Las actividades estructuradas pueden contener otras actividades y definen la lógica de negocio. A las actividades pueden asociarse un conjunto de atributos y un conjunto de contenedores. Estos últimos pueden incluir diferentes elementos, que a su vez pueden tener atributos asociados. En el listado 1.1 podemos ver un ejemplo explicativo.

Listado 1.1: Ejemplo de WS-BPEL

```
<flow>  ← Actividad estructurada
  <links>  ← Contenedor
    <link name="comprobarVuelo-reservarVuelo" ←Atributo/> ←Elemento
  </links>
  <invoke name="comprobarVuelo" . . . > ←Actividad basica
    <sources>  ←  Contenedor
      <source linkName="comprobarVuelo-reservarVuelo" ←Atributo/> ←Elemento
        </sources>
  </invoke>
  <invoke name="comprobarHotel" . . . />
  <invoke name="comprobarAlquilerCoche" . . . />
  <invoke name="reservarVuelo" . . . >
  <targets>  ← Contenedor
    <target linkName="comprobarVuelo-reservarVuelo" /> ←Elemento
  </targets>
  </invoke>
</flow>
```

1.4.2. WSDL

WSDL (Web Services Description Language) es un formato XML utilizado para describir la interfaz pública de servicios Web. Una composición WS-BPEL normalmente reúne varios servicios Web (descritos con WSDL) en uno de nivel superior (la composición), también descrito con WSDL.

WSDL consta de un conjunto muy amplio de reglas y normas. En la práctica, no se suelen implementar todas: hacerlo sería muy complejo y costoso. La mayoría de las implementaciones actuales se centran en el subconjunto definido por la especificación WS-I Basic Profile 1.1 [6]. Este subconjunto permite conseguir una mayor interoperabilidad entre todas las implementaciones existentes.

Está basado en XML y describe la forma de comunicación, es decir, los requisitos del protocolo y los formatos de los mensajes necesarios para interactuar con los servicios listados. Las operaciones y mensajes que utilizan los servicios se describen en abstracto y se ligan después al protocolo concreto de red y al formato del mensaje. Así, WSDL se usa a menudo en combinación con SOAP (§ 1.4.3) y XML Schema (§ 1.4.4).

Un programa cliente que se conecta a un servicio web puede leer el WSDL para determinar qué funciones están disponibles en el servidor. Los tipos de datos se describen en el archivo WSDL usando XML Schema. El cliente puede usar distintas tecnologías, en el grupo UCASE la elegida es SOAP para hacer la llamada a una de las funciones listadas en el WSDL. El WSDL nos permite tener una descripción de un servicio web. Especifica la interfaz abstracta a través de la cual un cliente puede acceder al servicio y los detalles de cómo se debe utilizar.

Podemos ver un ejemplo en el listado 1.2 donde la estructura WSDL tiene los siguientes elementos:

- Tipos de datos `<types>`: Esta sección define los tipos de datos usados en los mensajes. Se utilizan los tipos definidos en la especificación de esquemas XML.

- Mensajes `<message>`: Aquí definimos los elementos de mensaje. Cada mensaje puede consistir en una serie de partes lógicas. Las partes pueden ser de cualquiera de

los tipos definidos en la sección anterior.

- Tipos de puerto <portType>: Con este apartado definimos las operaciones permitidas y los mensajes intercambiados en el servicio.

- «Bindings» <binding>: Especificamos las definiciones de los protocolos de comunicación usados entre los servicio. El elemento binding está formado por dos atributos name y type, el primer define el nombre de la unión y el segundo especifica el protocolo usado.

- Servicios <service>: Conjunto de puertos y dirección de los mismos. Esta parte final hace referencia a lo aportado por las secciones anteriores.

Con estos elementos no sabemos qué hace un servicio pero sí disponemos de la información necesaria para interactuar con él.

Listado 1.2: Ejemplo de WSDL

```
<definitions name="StockQuote"
         targetNamespace="http://example.com/stockquote.wsdl"
         xmlns:tns="http://example.com/stockquote.wsdl"
         xmlns:xsd1="http://example.com/stockquote.xsd"
         xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
         xmlns="http://schemas.xmlsoap.org/wsdl/">

    <types>
       <schema targetNamespace="http://example.com/stockquote.xsd"
              xmlns="http://www.w3.org/2000/10/XMLSchema">
          <element name="TradePriceRequest">
             <complexType>
                 <all>
                     <element name="tickerSymbol" type="string"/>
                 </all>
             </complexType>
          </element>
```

```
18          <element name="TradePrice">
19             <complexType>
20                 <all>
21                     <element name="price" type="float"/>
22                 </all>
23             </complexType>
24          </element>
25       </schema>
26    </types>
27
28    <message name="GetLastTradePriceInput">
29        <part name="body" element="xsd1:TradePriceRequest"/>
30    </message>
31
32    <message name="GetLastTradePriceOutput">
33        <part name="body" element="xsd1:TradePrice"/>
34    </message>
35
36    <portType name="StockQuotePortType">
37        <operation name="GetLastTradePrice">
38           <input message="tns:GetLastTradePriceInput"/>
39           <output message="tns:GetLastTradePriceOutput"/>
40        </operation>
41    </portType>
42
43    <binding name="StockQuoteSoapBinding" type="tns:StockQuotePortType">
44        <soap:binding style="document"
45                              transport="http://schemas.xmlsoap.org/soap/http"/>
46        <operation name="GetLastTradePrice">
47           <soap:operation soapAction="http://example.com/GetLastTradePrice"/>
48           <input>
49               <soap:body use="literal"/>
50           </input>
```

```
51          <output>
52              <soap:body use="literal"/>
53          </output>
54       </operation>
55    </binding>
56
57    <service name="StockQuoteService">
58       <documentation>My first service</documentation>
59       <port name="StockQuotePort" binding="tns:StockQuoteSoapBinding">
60          <soap:address location="http://example.com/stockquote"/>
61       </port>
62       </service>
63
64 </definitions>
```

1.4.3. SOAP

SOAP (Simple Object Access Protocol) es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos XML. Este protocolo deriva de un protocolo creado por David Winer en 1998, llamado XML-RPC. SOAP fue creado por Microsoft, IBM y otros y está actualmente bajo el auspicio de la W3C. Es uno de los protocolos utilizados en los servicios Web. En el listado 1.3 podemos ver el esqueleto de un mensaje SOAP.

1.4.4. XML Schema

XML Schema es un lenguaje utilizado para describir la estructura y las restricciones de los contenidos de los documentos XML de una forma muy precisa, más allá de las normas sintácticas impuestas por XML. Fue desarrollado por el World Wide Web Consortium (W3C) y alcanzó el nivel de recomendación en mayo de 2001. Podemos ver un ejemplo en el listado 1.4 donde podemos ver que está compuesto por un elemento llamado «note» que a su vez está definido por cuatro elementos de tipo cadena denominadas

Listado 1.3: Ejemplo de SOAP

```
<?xml version="1.0"?>
<soap:Envelope
xmlns:soap="http://www.w3.org/2001/12/soap-envelope"
soap:encodingStyle="http://www.w3.org/2001/12/soap-encoding">

        <soap:Header>
                ...
        </soap:Header>

        <soap:Body>
                ...
                <soap:Fault>
                        ...
                </soap:Fault>
        </soap:Body>

</soap:Envelope>
```

Listado 1.4: Ejemplo de XML Schema

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="note">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="to" type="xs:string"/>
      <xs:element name="from" type="xs:string"/>
      <xs:element name="heading" type="xs:string"/>
      <xs:element name="body" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>

</xs:schema>
```

«to», «from», «heading» y «body».

1.4.5. BPELUnit

BPELUnit es una biblioteca de pruebas unitarias para composiciones WS-BPEL, creada por Philip Mayer. Puede usar cualquier motor que implemente WS-BPEL 2.0. Entre sus principales características está el uso de un formato XML (BPELUnit Test Specification o BPTS) para describir los casos de prueba a ejecutar y la posibilidad de sustituir servicios externos con otros servicios («mockups») que los simulen desarrollando el comportamiento indicado en la especificación proporcionada por el usuario. Además, BPELUnit ofrece posibilidades para añadir envíos síncronos y asíncronos. Podemos ver un ejemplo en el listado 1.5.

A partir de la versión 1.5 de *BPELUnit*, los ficheros BPTS incorporan el lenguaje de plantillas Apache Velocity. Las plantillas permiten generar los mensajes a partir de una

fuente de datos y una serie de variables predefinidas. Esto permite que el usuario pueda definir fácilmente diversos casos de prueba que tengan las mismas actividades, pero distinto contenido en los mensajes. Con ello, se obtienen, principalmente, tres ventajas:

- Facilita la generación de los casos de prueba y se hace más sencilla su automatización.
- Se ha separado la generación de casos de prueba de los detalles de WSDL y SOAP.
- Permite la creación de «mockups» más inteligentes, que consideren los mensajes que reciben (qué hay en la petición).

Listado 1.5: Ejemplo de un fichero BPTS

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <tes:testSuite
3    xmlns:esq="http://xml.netbeans.org/schema/Loans"
4    xmlns:ap="http://j2ee.netbeans.org/wsdl/ApprovalService"
5    xmlns:as="http://j2ee.netbeans.org/wsdl/AssessorService"
6    xmlns:sp="http://j2ee.netbeans.org/wsdl/LoanService"
7    xmlns:pr="http://enterprise.netbeans.org/bpel/N6 ServicioPrestamo/
8              LoanApprovalProcess"
9    xmlns:tes="http://www.bpelunit.org/schema/testSuite">
10
11   <tes:name>LoanServiceTest</tes:name>
12   <tes:baseURL>http://localhost:7777/ws</tes:baseURL>
13
14   <tes:deployment>
15     <tes:put name="LoanApprovalProcess" type="activebpel">
16       <tes:wsdl>LoanService.wsdl</tes:wsdl>
17       <tes:property name="BPRFile">LoanApprovalDoc.bpr</tes:property>
18     </tes:put>
19     <tes:partner name="assessor" wsdl="AssessorService.wsdl"/>
20     <tes:partner name="approver" wsdl="ApprovalService.wsdl"/>
```

```
21  </tes:deployment>
22
23  <tes:testCases>
24    <tes:testCase name="LargeAmount" basedOn="" abstract="false" vary="false">
25      <tes:clientTrack>
26        <tes:sendReceive
27           service="sp:LoanServiceService"
28           port="LoanServicePort"
29           operation="grantLoan">
30
31          <tes:send fault="false">
32            <tes:data>
33              <esq:ApprovalRequest>
34                <esq:amount>150000</esq:amount>
35              </esq:ApprovalRequest>
36            </tes:data>
37          </tes:send>
38
39          <tes:receive fault="false">
40            <tes:condition>
41              <tes:expression>esq:ApprovalResponse/esq:accept</tes:expression>
42              <tes:value>'true'</tes:value>
43            </tes:condition>
44          </tes:receive>
45
46        </tes:sendReceive>
47      </tes:clientTrack>
48
49      <tes:partnerTrack name="approver">
50        <tes:receiveSend
51           service="ap:ApprovalServiceService"
52           port="ApprovalServicePort"
53           operation="approveLoan">
```

```
54
55          <tes:send fault="false">
56            <tes:data>
57              <esq:ApprovalResponse>
58                <esq:accept>true</esq:accept>
59              </esq:ApprovalResponse>
60            </tes:data>
61          </tes:send>
62
63          <tes:receive fault="false"/>
64        </tes:receiveSend>
65      </tes:partnerTrack>
66
67      <tes:partnerTrack name="assessor"/>
68    </tes:testCase>
69    ...
70  </tes:testCases>
71 </tes:testSuite>
```

En el listado 1.5 podemos observar la estructura de un fichero BPTS mediante un fragmento de un fichero de pruebas de la clásica composición WS-BPEL de aprobación de un préstamo («LoanApproval»). Esta composición recibe un mensaje de un cliente que solicita una cierta cantidad de dinero. Dependiendo de la cantidad solicitada, el proceso WS-BPEL invoca al WS asesor («asessor») cuando la cantidad que se solicita es menor o igual a 10000, o al WS aprobador («approver»), en otro caso. La salida del WS asesor se corresponde con el nivel de riesgo del cliente. Si el riesgo es bajo, se concede el préstamo, en caso contrario la petición se envía al WS aprobador, el cuál toma la decisión final de aceptación del préstamo.

Hasta la línea 8 tenemos el elemento raíz y la definición de los espacios de nombres XML. En este caso, el prefijo `tes` está asociado al espacio de nombres de BPELUnit y el resto, son prefijos propios de esta composición y el de los envoltorios SOAP.

A continuación encontramos la sección de despliegue (líneas 10–20). En ella podemos

comprobar que la composición se comunica con dos «mockups»: el asesor y el aprobador.

Finalmente, de la línea 22 en adelante, aparece la sección de los casos de prueba en la que debe ir incluido cada bloque `testCase`. En concreto, se ha incluido un caso de prueba de ejemplo en el que el cliente solicita un préstamo por valor de 150.000€ (líneas 30–36), que es aceptado directamente por el aprobador (líneas 52–58), sin que tenga que intervenir el asesor (línea 63). Finalmente, el aprobador comunica al cliente su decisión (líneas 38–44).

1.4.6. Apache Velocity

Apache Velocity es un motor de plantillas basado en Java. Le permite a los diseñadores de páginas hacer referencia a métodos definidos dentro del código Java. Los diseñadores Web pueden trabajar en paralelo con los programadores Java para desarrollar sitios de acuerdo al modelo de Modelo-Vista-Controlador (MVC), permitiendo que los diseñadores se concentren únicamente en crear un sitio bien diseñado y que los programadores se encarguen solamente de escribir código de primera calidad. Velocity separa el código Java de las páginas Web, haciendo el sitio más mantenible a largo plazo y presentando una alternativa viable a Java Server Pages (JSP) o PHP.

BPELUnit utiliza Velocity de dos formas: como lenguaje de plantillas propiamente dicho para generar los mensajes a enviar a la composición WS-BPEL, y como formato de entrada para recibir los valores de las variables a usar en las plantillas de los mensajes. Este PFC se centra en el segundo uso.

En el listado 1.6 podemos ver un ejemplo sencillo de un fichero *html* con código Velocity, dónde podemos observar que existe una variable para definir el contenido de `h1` y un bucle para crear una tabla con los alumnos. Si dicho fichero se ejecutara asignando “Tabla de estudiantes” a `titulo`, “white” a `color` y una lista con “Pedro”, “Ana” y “Juan” a `listado`, la salida resultante de dicho ejemplo es la mostrada en 1.7.

Para que quede más claro, también vamos a mostrar un ejemplo más parecido a los que va a tratar este PFC. Puede verse un ejemplo de plantilla Velocity en el listado 1.8.

Los ficheros Velocity están formados por filas, cada fila representa los valores de cada

Listado 1.6: Ejemplo de una plantilla Velocity

```
<html>
       <body>
              <h1 > $titulo </h1>
              <br/>
              tabla
              <table>
              #foreach ($alumno in $listado)
                     <tr><td bgcolor="$color">$alumno</td></tr>
              #end
              <table/>
       </body>
</html>
```

Listado 1.7: Salida producida por la plantilla Velocity

```
<html>
       <body>
              <h1 > Tabla de Estudiantes </h1>
              <br/>
              tabla:
              <br/>
              <table/>
                     <tr><td bgcolor="white">Pedro</td></tr>
                     <tr><td bgcolor="white">Juan</td></tr>
                     <tr><td bgcolor="white">Ana</td></tr>
              <table/>
              <br/>
   </body>
</html>
```

una de las variables para cada caso de prueba. La primera componente corresponderá al primer caso de prueba, la segunda al segundo, etc.

Las variables que se vayan a declarar en la fuente de datos cuando el tipo es Velocity obligatoriamente han de listarse (separadas por espacios) en la propiedad iteratedVars. Todas las variables que no se declaren aquí pero sí aparezcan en la fuente de datos, únicamente se copiaran tal cual, pero no se sustituirán en las plantillas.

Todas las variables incluidas en `iteratedVars` deben tener asociadas en la fuente de datos una lista de valores con idéntico número de elementos.

Listado 1.8: Ejemplo de plantilla Velocity para generar un mensaje SOAP

```
<bpelunit:send fault="false">
  <bpelunit:template>
    <srv:ApprovalResponse>
      #set( $sAmount = $xpath.evaluateAsString("//srv:amount", $request) )
      #set( $amount = $integer.parseInt($sAmount) )
      #if( $ap_reply != 'smart' )
      <srv:accept>$ap_reply</srv:accept>
      #elseif( $ap_limit > $amount )
      <srv:accept>true</srv:accept>
      #else
      <srv:accept>false</srv:accept>
      #end
    </srv:ApprovalResponse>
  </bpelunit:template>
</bpelunit:send>
```

1.4.7. CSV

CSV («comma-separated values» o «valores separados por comas») describe a un conjunto de formatos de fichero basados en texto plano donde cada línea representa una fila, y los valores de cada fila están separados por un marcador determinado. Normalmente, este marcador es una coma (de ahí su nombre).

CSV es otro formato de entrada que permite *BPELUnit* para sus plantillas. Puede verse un ejemplo en el listado 1.9. Cada fila proporciona los datos de un trayecto de tren: tipo de máquina, origen, destino, número de trenes y después las horas de salidas.

Listado 1.9: Ejemplo de un fichero CSV

```
tren,Origen,Destino,númeroTrenes,horasSalidas
pequenio,cádiz,sevilla,1,9:00,11:00,13:00,15:00
grande,cádiz,sevilla,1,17:00,19:00,19:00,21:00
grande,sevilla,huelva,1,9:00,11:00
ave,cádiz,sevilla,1,9:15,10:15
pequenio,sevilla,jerez,1,9:00,10:00,13:00,14:00
```

Capítulo 2

Planificación

En este capítulo veremos cómo se ha organizado temporalmente el trabajo de este proyecto así como las tareas que han tenido lugar y la planificación de las mismas.

El proyecto se ha desarrollado a lo largo de unos diez meses de trabajo, empezando por febrero del 2011 hasta noviembre del 2011. No obstante hasta mediados de junio no ha tenido una ocupación plena del tiempo, puesto que anteriormente tenía otro tipo de ocupaciones incompatibles con la dedicación íntegra del proyecto.

La idea del proyecto empezó unos meses antes: fue cuando decidí ponerme en contacto con la profesora Inmaculada Medina Bulo, coordinadora del grupo de investigación UCASE. Ella me invitó a asistir a unos de los seminarios del grupo que realizan cada semana en el cual me presentó al profesor Antonio García Domínguez, quien me propuso realizar un generador de datos aleatorios ya que el grupo de investigación necesitaba de dicha herramienta. Así surgió *TestGenerator*, pero por motivos profesionales y académicos la idea no empezó a tomar forma hasta febrero del 2011.

2.1. Etapas

La metodología usada para el modelo del software ha sido un desarrollo iterativo por prototipos. En la etapa de inicialización del desarrollo iterativo, se realizó una versión completa del programa pero trataba un subconjunto de datos y restricciones además

de que el formateador sólo era capaz de hacerlo en formato Apache Velocity. En las siguientes iteraciones se fueron añadiendo las demás funcionalidades del sistema.

El desarrollo seguido se puede dividir en distintos apartados que se especificarán en los siguientes puntos.

2.1.1. Elicitación de requisitos

Los requisitos fueron analizándose a través de varias reuniones con los miembros del grupo de investigación, aunque fueron afinándose con posteriores reuniones con el profesor Antonio García Dominguez. Los requisitos fueron incrementándose a lo largo del proyecto, por ejemplo el requisito de diseñar una gramática tuvo lugar a posteriori.

2.1.2. Estudio de tecnologías

En esta fase se ha llevado el estudio de varias tecnologías en su mayoría prácticamente desde cero. Algunas de ellas venían impuestas por el grupo y otras se fueron eligiendo a medida que se desarrollaba. Entre las tecnologías usadas cabe destacar el lenguaje de programación Java usando de framework NetBeans, *JUnit* para las pruebas unitarias, *SVN* para el control de versiones, *XMLBeans* para recorrer archivos XML, *ANTLRWorks* para generar el AST, LaTeX para realizar la documentación y *Maven* para gestión y construcción del proyecto.

2.1.3. Analizador de plantillas de *ServiceAnalyzer*

En este apartado se llevó a cabo el estudio de cómo recorrer de manera cómoda los catálogos de *ServiceAnalyzer* con el fin de poder quedarnos con la parte que nos interesa y posteriormente poder crear las estructuras de datos para guardar la información útil para nuestro proyecto.

2.1.4. Creación de estructuras de tipos

En este punto se estudió e implementó una estructura de datos capaz de representar no sólo los datos que posteriormente íbamos a generar de forma aleatoria sino también las restricciones que puedan tener. Además de esto, dicha estructura tenía que ser lo suficientemente sólida e independiente del analizador y el formateador para que si la aplicación creciera dicha estructura siguiese siendo válida, cosa que ocurrió al implementarse el analizador TestSpec y el formateador CSV, puesto que en un principio esto no era un requisito.

2.1.5. Generador de datos aleatorios

Esta es la parte más importante del sistema, se podría decir que es la corazón del mismo. Una vez creada las estructuras de tipos, entraba en juego la generación de los datos de forma aleatoria y uniformemente distribuida. Creándose como resultado los datos que luego van a ser formateados para la salida del sistema.

2.1.6. Creación de gramática

Para que la aplicación no estuviese acotada a los archivos *wsdl*, se creo un lenguaje capaz de representar los tipos de datos que maneja el sistema con sus restricciones. Se llamó TestSpec y sus archivos tienen la extensión *spec*.

2.1.7. Analizador del lenguaje TestSpec

Para poder tomar los datos de dicha gramática, fue necesaria la implementación de un analizador que generase el AST (árbol de sintaxis abstracta) apropiado con la ayuda de la herramienta *ANTLRWorks*.

2.1.8. Formateador Apache Velocity

Una vez que teníamos los datos generados de forma aleatoria, era necesario formatearlos para la creación de ficheros que luego sirviesen de entrada en *BPELUnit* para las

pruebas paramétricas.

2.1.9. Formateador CSV

Dicho formato también es aceptado para las pruebas paramétricas de *BPELUnit*. Dicho formato es más sencillo de usar, pero menos potente.

2.1.10. Pruebas unitarias

Cada parte del código lleva asociada unas pruebas unitarias gracias a *JUnit*, esto nos permite detectar errores y poder subsanarlos. Aparte de esto, nos permite que al incluir una nueva modificación al programa poder comprobar si sigue comportándose de la forma deseada. Esta parte ha llevado un gran esfuerzo para que la aplicación sea sólida y consistente.

2.1.11. Documentación

La memoria del este proyecto se ha ido elaborando conforme se iba desarrollando la aplicación, pero es en los últimos meses donde se ha trabajado más esta parte.

2.1.12. Paquete de instalación y líneas de ordenes

En esta fase se llevaron a cabo los ajustes necesarios para que *TestGenerator* fuera fácil de instalar y usar desde una consola a través de líneas de órdenes.

2.2. Diagrama Gantt

Se ha elaborado un diagrama Gantt para facilitar la comprensión de la distribución de las tareas. Para elaborar dicho programa, se ha usado el software *Gantt Project*. El diagrama se muestra en las figuras 2.1 y 2.2.

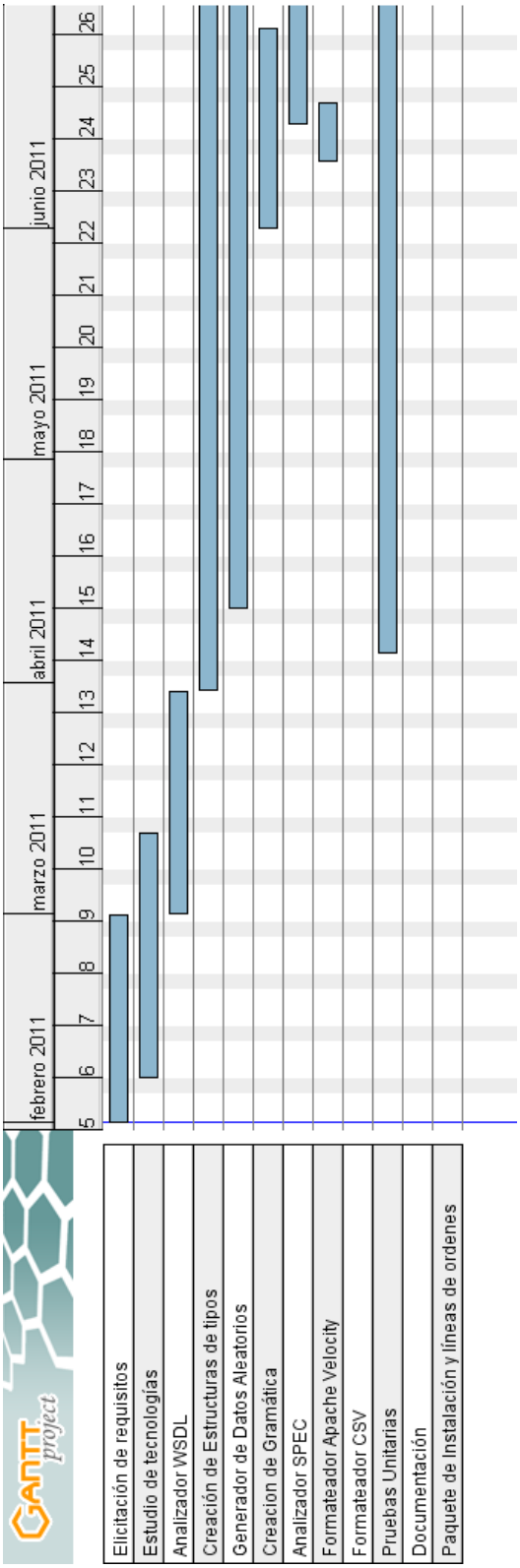

Figura 2.1.: Diagrama de Gantt I

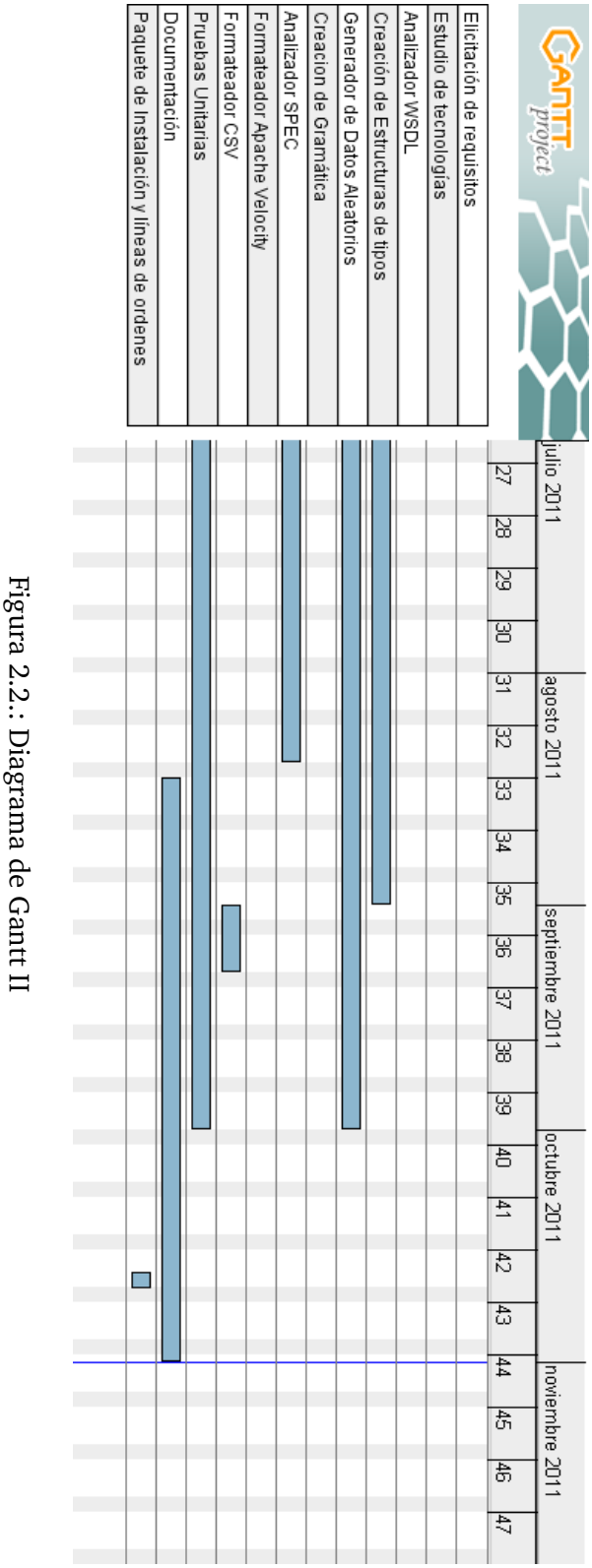

Figura 2.2.: Diagrama de Gantt II

Capítulo 3

Análisis

En este apartado hablaremos de las necesidades que debe cubrir el proyecto, es decir, los requisitos que debe cumplir. Además se realizarán los casos de uso que especifican el alcance del proyecto.

3.1. Requisitos funcionales

En este punto hablaremos de las necesidades que debe cubrir el sistema. Dichas necesidades vienen impuestas por el grupo UCASE las cuales veremos a continuación.

TestGenerator producirá los datos que necesitan las plantillas que produce *ServiceAnalyzer* de forma aleatoria, y usaremos su diseño general como punto de partida y para validar estrategias de generación futuras. Puede operar directamente con las plantillas de *ServiceAnalyzer* si hay que trabajar con un mensaje suelto, o con una declaración de variables y tipos si tenemos que generar un fichero de datos para un conjunto parametrizado de casos de prueba de *BPELUnit*. Para el objetivo anterior y también para obtener una mayor flexibilidad y que el uso de la herramienta no quede acotado a las plantillas generadas por *ServiceAnalyzer*, el programa deberá también aceptar como entrada unos ficheros escritos en un lenguaje de dominio específico que sea capaz de representar este conjunto de datos y restricciones.

El sistema deberá ser capaz de generar datos aleatorios con las restricciones especi-

ficadas en las plantillas que produce *ServiceAnalyzer* y generar de salida un fichero con dichos datos para poder integrarlos en BPELUnit.

El conjunto de datos que debe ser capaz de generar de forma aleatoria es el siguiente:

- Enteros.
- Números con coma flotante.
- Cadenas de caracteres.
- Fechas.
- Horas.
- Fechas y horas.
- Duraciones.
- Listas.
- N-Tuplas.

Bajo las siguientes restricciones:

- Valor máximo.
- Valor mínimo.
- Número total de dígitos.
- Número total de decimales.
- Cumplir cierta expresión regular.
- Valor válido dada una lista de valores.

Para poder llevar a cabo un control de los datos a generar y sus restricciones, será necesario realizar unas estructuras o tipos de datos propios donde se guardará la información relevante de los mismos. Este sistema de tipos servirá para que la herramienta pueda comunicarse en cada una de sus fases y para poder modular la aplicación.

Además, el sistema deberá ser capaz de exportar los datos que genera en formatos compatibles con *BPEUnit*, para ello se propone los siguientes formatos:

- CSV: la primera tupla representará el nombre de las variables y cada tupla posterior, definirá un caso de prueba.
- Apache Velocity: la forma en que representaremos las variables y sus valores de caso de prueba será la siguiente `#set($NombreVariable = [valor1, valor2])` Donde `valor1`, `valor2`... representará cada uno el valor de la variable para un determinado caso de prueba.

Aunque CSV tiene la gran de ventaja de ser muy sencillo y fácil de usar, pierde en potencia: por ejemplo, no es capaz de representar una lista de elementos. Por ello, es necesario el uso del formato Apache Velocity. Es un formato más complejo que el anterior, pero mucho más potente.

3.2. Estudio del mercado

Antes de empezar a desarrollar este PFC, se ha llevado a cabo un estudio de mercado para ver que herramientas existían actualmente para Java.

Básicamente se encontraron dos herramientas que se dedica a generar datos aleatorios estas son:

- *JCheck* [8]
- *QuickCheck for Java* [9]

Dichas herramientas se centran exclusivamente en dar un API para Java, de cara a ser integrados con *JUnit*. En el listado 3.1 podemos ver un ejemplo de uso de *JCheck*. El ejemplo de *QuickCheck for Java* lo veremos en el listado 3.2

Se podría decir que la mayor diferencia que tiene con nuestra aplicación es el enfoque; *TestGenerator* está pensado para generar ficheros de datos genéricos a introducir en cualquier herramienta como por ejemplo *BPELUnit*, que es la que usa el grupo de

Listado 3.1: Ejemplo de uso de *JCheck*

```
@RunWith(org.jcheck.runners.JCheckRunner.class)
class SimpleTest {
       @Test public void simpleAdd(int i, int j) {
       Money miCHF= new Money(i, "CHF");
       Money mjCHF= new Money(j, "CHF");
       Money expected= new Money(i+j, "CHF");
       Money result= miCHF.add(mjCHF);
       assertTrue(expected.equals(result));
    }
}

@RunWith(org.jcheck.runners.JCheckRunner.class)
class SimpleTest {
       @Test public void simpleDivide(int[] somearray) {
       imply(somearray.length > 1);
       int expected= somearray.length;
       int result= MySort.sort(somearray).length;
       assertEquals(expected, result);
    }
}
```

Listado 3.2: Ejemplo de uso de *QuickCheck for Java*

```
public class SortedListTest {
	@Test public void sortedListCreation() {
		for (List<Integer> any : someLists(integers())) {
			SortedList sortedList = new SortedList(any);
			List<Integer> expected = sort(any);
			assertEquals(expected, sortedList.toList());
		}
	}

	private List<Integer> sort(List<Integer> any) {
		ArrayList<Integer> sorted = new ArrayList<Integer>(any);
		Collections.sort(sorted);
		return sorted;
	}
}
```

investigación UCASE. Otros programas se centran exclusivamente en dar un API para Java, de cara a ser integrados en JUnit.

Aparte de esta gran diferencia de enfoque, *TestGenerator* tendrá soporte para crear fechas aleatorias, no contempladas ni por *JCheck* ni por *QuickCheck for Java*.

3.3. Requisitos de implementación

Los requisitos de implementación impuestos por el grupo UCASE son los siguientes:

- El lenguaje a utilizar para la implementación de la solución deberá ser Java [10], puesto que la mayoría de las aplicaciones del grupo están escritas en dicho lenguaje y así resultará más cómodo una posible reutilización de código para otros proyectos.

- También como requisitos indispensable, el grupo UCASE exige la realización de pruebas paramétricas mediante la herramienta JUnit [11] §5.4.

- El grupo de investigación UCASE exige para realizar todos sus proyectos, utilizar un sistema de integración continua para poder llevar un mejor seguimiento de los mismos §5.1.

3.4. Atributos del sistema

El sistema debe tener:

- Mantenibilidad: el sistema deberá ser cómodo de mantener, dado que en un futuro será ampliado.

- Transportabilidad: el sistema deberá tener independencia de la plataforma, para no ligar su uso a ningún sistema operativo. Aunque se recomienda el uso de alguna distribución GNU/Linux.

- Facilidad de uso: aunque este proyecto está dirigido a personas con conocimientos informáticos elevados, deberá tener facilidad de uso para que no resulte una tarea tediosa trabajar con la aplicación.
- Licencia libre: dado que se desea un uso público de la misma, la mejor manera de conseguir es otorgándole dicha licencia.

3.5. Casos de uso

En la figura 3.1 podemos ver representado el diagrama de casos de uso, usando la notación UML. En las siguientes subsecciones detallaremos los contenidos de cada uno.

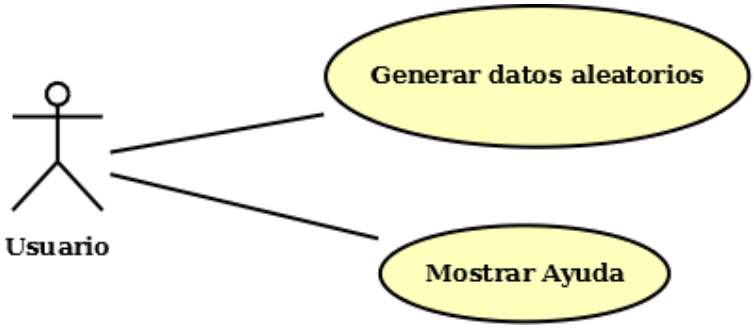

Figura 3.1.: Diagrama de casos de uso

3.5.1. Generar datos aleatorios

- Actor principal: Usuario que desea generar datos aleatorios.
- Precondición: Debe tener un fichero válido de entrada al programa.
- Postcondición: Genera un archivo que contendrá los datos generados aleatoriamente.
- Escenario principal:
  1. El usuario introduce como argumento la ruta de una plantilla generada por *ServiceAnalyzer*, el número de datos a generar y el formato de salida.

2. El sistema genera los datos en el formato especificado.

- Variaciones:

1a. La orden introducida no es correcta

1. El sistema informa que la orden no es válida, muestra su forma de uso y cancela el caso de uso.

1b. El número de argumentos introducidos por la línea de ordenes es incorrecto.

1. El sistema informa que el número de argumento no es válido, muestra su forma de uso y cancela el caso de uso.

1c. El usuario no introduce el número de datos a generar.

1. El sistema utiliza el valor por defecto y continua con el caso de uso.

1d. El usuario no introduce el formato en el que desea la salida.

1. El sistema utiliza el valor por defecto y continua con el caso de uso.

3.5.2. Mostrar ayuda

- Actor principal: Usuario que desea ver la ayuda.
- Precondición: Ninguna.
- Postcondición: Desde la terminal, se visualiza la ayuda del sistema.
- Escenario principal:

1. El usuario pide al sistema que le muestre la ayuda.
2. El sistema muestra la ayuda por la pantalla.

Capítulo 4

Diseño

En este capítulo hablaremos de las condiciones que nos llevarán a tomar las decisiones de diseño, así como que decisiones se ha llevado a cabo y de la estructura que tendrá el proyecto.

4.1. Arquitectura del sistema

En este punto, se dará a conocer los aspectos relativos de cómo será la estructura de este sistema.

En este proyecto, se ha seguido el patrón arquitectónico «Pipes and Filters» (tuberías y filtros) [12]. Dicho patrón arquitectónico proporciona una estructura para sistemas que procesan un flujo de datos. Cada paso de procesamiento se encapsula en un componente de filtro, los datos se pasan a través de tuberías entre los filtros adyacentes. La recombinación de filtros, permite crear sistemas relacionados.

Este proyecto, se encuentra dividido en tres grandes bloques (filtros). Esta división está basada en la funcionalidad que cumplen cada uno de los bloques. Los bloques son:

- Analizador
- Generador
- Formateador

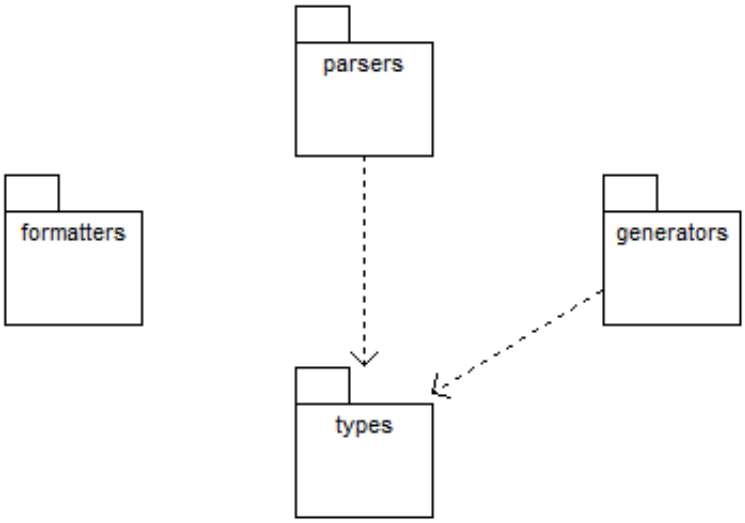

Figura 4.1.: Diagrama de los paquetes

La división en estos tres bloques, nos aportan grandes ventajas entre ella podemos destacar:

- Desarrollo: con esta estructura, es más sencillo y cómodo usar la metodología llevada en este proyecto que ha sido iterativa por prototipo.
- Corrección de errores: al estar claramente diferenciadas las funcionalidades, es más cómodo identificar problemas y corregir errores, asegurando que el resto de partes funcionen correctamente.
- Ampliación del sistema: si en algún momento necesitamos ampliar algún formateado extra o ampliar los tipos de ficheros de entrada al programa, gracias a dicha división es menos costoso y reduce la aparición de errores.
- Reutilización de código: puesto que el proyecto pertenece al grupo de investigación UCASE, es más sencillo la reutilización de alguna de sus partes en otros proyecto, cosa que ya se está pensando en hacer.

En la figura 4.1 podemos ver cómo se comunican los paquetes entre sí.

En apartados posteriores, vamos a definir cada uno de los bloques mencionados (Analizador, Generador y Formateador).

4.2. Formatos de ficheros de entrada

Aunque nuestra aplicación lo que reciba sea un fichero *wsdl*, no es con este directamente con el cual trabajaremos. En su lugar, nuestro programa procesará con ayuda de *ServiceAnalyzer* dicho fichero y generará unos catálogos con lo que trabajará internamente.

En el listado 4.1 veremos un ejemplo de un catálogo generado por *ServiceAnalyzer* con el cual podría trabajar nuestra aplicación.

Listado 4.1: Ejemplo de un catálogo de *ServiceAnalyzer*

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <mes:services xmlns:mes="http://serviceAnalyzer/messageCatalog">
3    <mes:service name="squaresSumService"
4                          uri="http://j2ee.netbeans.org/wsdl/squaresSum">
5      <mes:port name="squaresSumPort">
6        <mes:operation name="squaresSumOperation">
7          <mes:input>
8            <mes:decls>
9              <mes:typedef name="T1" type="int" min="0" max="4294967295"/>
10             <mes:variable name="squaresSumRequest" type="T1"/>
11           </mes:decls>
12           <mes:template><![CDATA[
13           <ns1:squaresSumRequest
14           xmlns:ns1="http://xml.netbeans.org/schema/squaresSum">
15           <ns1:n xmlns:ns1="http://xml.netbeans.org/schema/squaresSum">
16           $squaresSumRequest
17           </ns1:n>
18           </ns1:squaresSumRequest>]]>
19           </mes:template>
```

```
20        </mes:input>
21        <mes:output>
22          <mes:decls>
23            <mes:typedef name="T1" type="int" min="0" max="4294967295"/>
24            <mes:typedef name="T2" type="list" element="T1" min="1"/>
25            <mes:typedef name="T3" type="tuple" element="T1, T2"/>
26            <mes:variable name="squaresSumResponse" type="T3"/>
27          </mes:decls>
28          <mes:template><![CDATA[
29            <ns1:squaresSumResponse
30              xmlns:ns1="http://xml.netbeans.org/schema/squaresSum">
31            <ns1:sum xmlns:ns1="http://xml.netbeans.org/schema/squaresSum">
32              $squaresSumResponse.get(0)
33            </ns1:sum>
34            <ns1:elements xmlns:ns1="http://xml.netbeans.org/schema/squaresSum">
35              #foreach($V1 in $squaresSumResponse.get(1))
36            <ns1:element xmlns:ns1="http://xml.netbeans.org/schema/squaresSum">
37              $V1
38            </ns1:element>
39            #end
40            </ns1:elements>
41            </ns1:squaresSumResponse>]]>
42          </mes:template>
43        </mes:output>
44      </mes:operation>
45    </mes:port>
46  </mes:service>
47 </mes:services>
```

A nuestra aplicación, se le deberá indicar que servicio, operación y tipo se van a tratar para luego generar los datos de forma aleatoria. En el caso de la plantilla de ejemplo que se le muestra se le podría especificar que el servicio es “squaresSumService”, la operación “squaresSumOperation” y el tipo es “input”. Entonces nuestra aplicación deberá generar

datos aleatorios de la variable “squaresSumRequest” donde esta será de tipo int con un valor mínimo igual a “0” y un valor máximo “4294967295”.

Otro tipo de archivos que podrá recibir son especificaciones en un lenguaje diseñado específicamente para este programa, que denominaremos TestSpec.

TestSpec es un lenguaje muy sencillo que nos servirá para poder declarar los tipos de datos y restricciones contemplados en *ServiceAnalyzer*. Estará dividido en dos bloques: en primer lugar, un bloque de definiciones de tipos con sus restricciones y un segundo lugar, un bloque con las declaraciones de dichos tipos. En la sección 4.4 veremos más detalles acerca de dicho lenguaje.

4.3. Sistema de tipos

El dominio de datos del sistema deberá ser equivalente al conjunto de datos y restricciones que puede tomar del catálogo generado por *ServiceAnalyzer*.

Dicho conjunto de datos es el siguiente:

- `string`: Cadena de caracteres válidos Unicode e ISO/IEC 10646.
- `int`: Enteros que pueden almacenarse en 32 bits.
- `float`: Representa números de coma flotante de 32 bits, según el estándar IEEE 754.
- `date`: Define un día concreto del calendario Gregoriano con el formato `YYYY-MM-DD`, como por ejemplo, “2003-10-21”.
- `time`: Define una hora concreta con el formato `hh:mm:ss`, como por ejemplo, “10:21:23”. El número de segundos puede incluir dígitos decimales de precisión arbitraria si se desea. Las horas se datan según el sistema de 24 horas.
- `dateTime`: Define un instante de tiempo concreto, usando el calendario Gregoriano. El formato es `YYYY-MM-DDThh:mm:ss`, por ejemplo, “2003-10-21T20:30:13”. La T separa la fecha de la hora. Se puede añadir también una Z opcional, y + o

- hh:mm al final para indicar una zona horaria diferente. Usamos Z si la hora se ajusta GMT (Greenwich Mean Time) o o a UTC (Coordinate Universal Time), o utilizaremos las horas y minutos adicionales para indicar diferencias respecto al GMT.

- `duration`: Expresa una duración en un espacio de 6 dimensiones. El formato es `PnYnMnDTnHnMnS`, donde `nY` representa el número de años, `nM` el número de meses, `nD` el número de días `nH` el número de horas, `nM` el número de minutos y `nS` el de segundos. `P` es un indicador obligatorio que debe estar el primero, mientras que T es el carácter que separa la fecha de la hora y únicamente debe aparecer si se indica una hora. Ninguno de los elementos es obligatorio ni tiene limitación de rango. También se pueden indicar duraciones negativas precediéndolas del signo '-' (si se omite el signo, se tratarán como duraciones positivas). Un valor de tipo duration podrá ser, por ejemplo, "P1DT2S" (duración de un día y dos segundos).

También se van a contemplar algunos tipos contenedores, en concreto las listas y las tuplas:

- `list`: Una lista es una colección de elementos que contiene el mismo tipo de dato.

- `tuple`: Una tupla es una colección de elementos los cuales pueden ser de distintos tipos.

El conjunto de restricciones que pueden aparecer en la plantilla es el siguiente:

- `element`: Esta restricción es obligatoria si el tipo es alguno de los contenedores, representa el tipo de elemento que va a contener. En el caso de las tuplas podríamos necesitar una lista de tipos, por lo que el valor de `element` es una lista ordenada de cadenas separadas por coma.

- `min`: Este atributo tiene un significado u otro en función del tipo al que se aplique. Aplicado a tipos numéricos, representa el límite inferior inclusivo del espacio de valores del tipo. En el caso de que se aplique a un tipo cadena, indica la longitud

mínima que ésta ha de tener. Sin embargo, si es un tipo `list`, `min` indica el número mínimo de elementos que puede contener la lista.

- `max`: Análogamente a la definición de `min` se le puede aplicar sustituyendo la palabra mínimo por máximo.
- `values`: Restringe el espacio de valores de un determinado tipo al conjunto de valores especificados.
- `pattern`: Restringe el espacio de valores de un determinado tipo, restringiendo el espacio léxico a literales que siguen un determinado patrón. El valor debe ser una expresión regular.
- `fractionDigits`: Controla el número de decimales que deberá contener `float`.
- `totalDigits`: Controla el número total de dígitos que tendrá un número.

En las figuras 4.2 y 4.3 podemos ver el diagrama de clases que representa cómo vamos a guardar los tipos y sus restricciones.

4.4. Analizadores

Este bloque estructural del proyecto, es el encargado de analizar los ficheros que recibe de entrada con el fin de crear una estructura de datos que guarde la información relevante para, en una etapa más tardía, poder generar los datos aleatorios. El sistema es capaz de analizar dos tipos de documentos, los catálogos generados por *ServiceAnalyzer* a partir de un fichero *wsdl* del grupo de investigación y ficheros de un lenguaje diseñado en este proyecto llamado TestSpec.

Para el análisis de los ficheros *wsdl* utilizaremos *ServiceAnalyzer*. Dicho programa nos generará un catálogo que será el que analizaremos utilizando las clases del propio *ServiceAnalyzer*, las cuales utilizan la herramienta *XMLBeans*. Gracias a dicha herramienta seremos capaz de recorrerlo, y a partir de esto, recoger la información en unas estructuras creadas para tal fin que contendrá las restricciones de los datos que debemos generar.

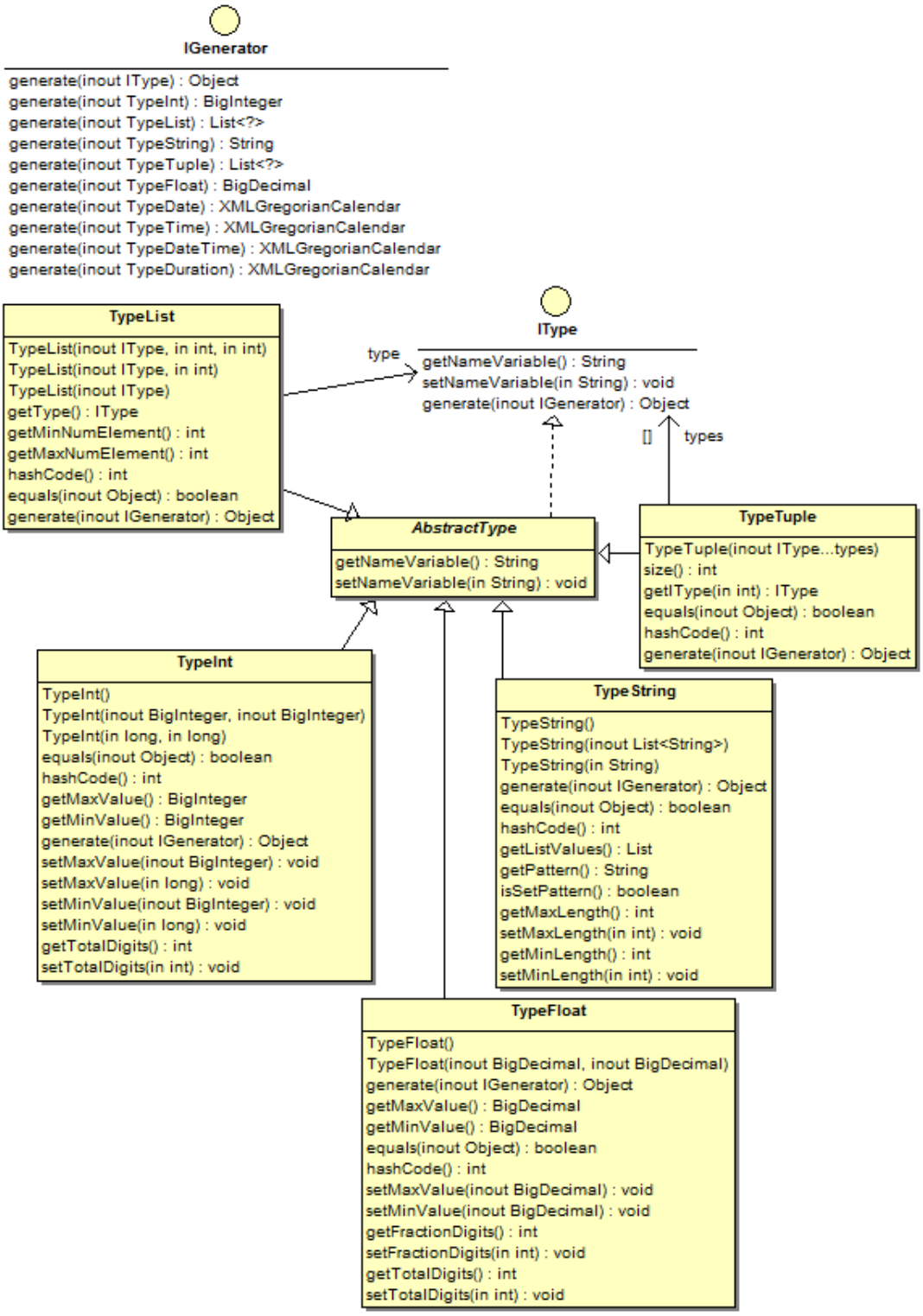

Figura 4.2.: Diagrama de los tipos I

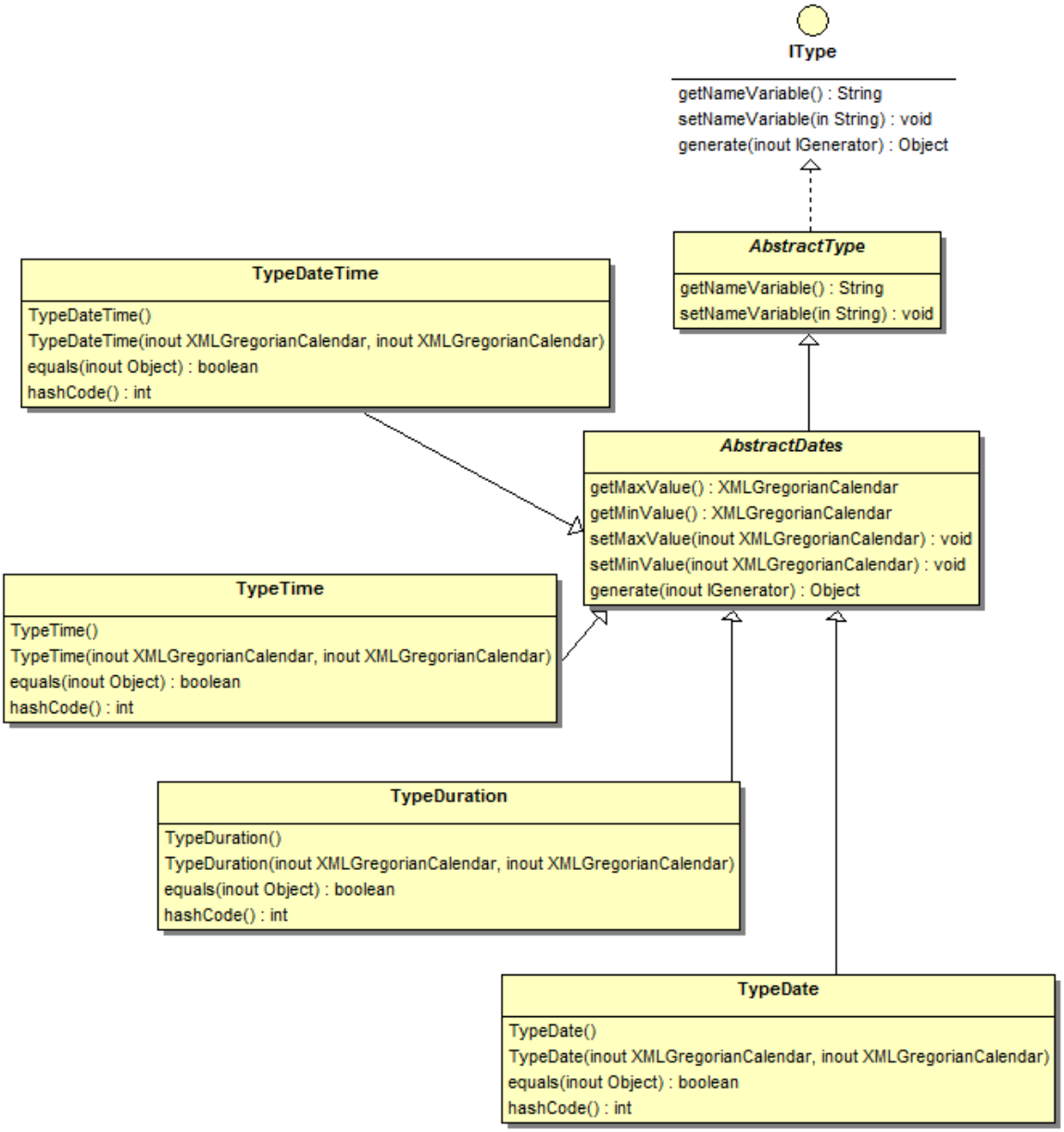

Figura 4.3.: Diagrama de los tipos II

Para los ficheros con extensión *spec*, se ha creado para este proyecto un lenguaje denominado TestSpec. A continuación pasamos a definir su estructura.

Una especificación TestSpec estará dividida en dos bloques: en el primero aparecerán las definiciones de los tipos con sus restricciones la cual denominaremos `typedef`, en el segundo y último aparecerán las declaraciones de los tipos el cual llamaremos `declaration`. Veamos un ejemplo.

Listado 4.2: Ejemplo de especificación TestSpec

```
1 typedef int (min=1, max=6) CaraDado;
2 CaraDado dado;
```

En el listado 4.2 podemos observar que las líneas que pertenecen al bloque de definiciones aparece la palabra reservada `typedef`, a continuación el tipo que vamos a definir (`int`) y después, aparecen el conjunto de restricciones formadas por los pares clave-valor (`min=1,max=6`). Para terminar, al final de la línea aparece el nombre que le vamos a dar a dicha definición (`CaraDado`) precedido del carácter fin de línea “;”.

A su vez un `typedef` puede servir de tipo para otro `typedef` posteriormente declarado.

En las sentencias de las declaraciones, aparecerá el nombre de una de las definiciones anteriores (`CaraDado`) y el identificador que le vamos a otorgar a la variable (`dado`) precedido del carácter fin de línea “;”.

Este tipo de estructura, hace de TestSpec un lenguaje muy cómodo y sencillo de usar para nuestro objeto: especificar un conjunto de restricciones asociadas a un tipo de dato.

Mostraremos un ejemplo algo más elaborado para que quede clara su potencia en el listado 4.3.

Listado 4.3: Ejemplo complejo de especificación TestSpec

```
1 typedef string (values={"true", "false"}) boolean;
2 typedef int (min=1, max=6) CaraDado;
3 typedef tuple (element = {string, caraDado}) ResultadoTirada;
4 typedef list (min=1, element = int) ListaNoVaciaInt;
5 typedef int (min=0) positiveNumber;
6 typedef positiveNumber (max=100) smallNumber;
```

```
7
8  boolean myBoolean;
9  ResultadoTirada tirada;
10 ListaNoVaciaInt numeros;
11 smallNumber myNumber;
```

Las palabras reservadas para especificar las restricciones son las siguientes:

- min
- max
- values
- totalDigits
- fractionDigits
- element
- pattern

Una vez que tenemos la especificación, hace falta generar un AST («Abstract Syntax Trees», en español Árboles de Sintaxis Abstracta) para poder recorrerla y ser capaz de guardar la información en nuestra estructura de datos propia. Dicho árbol estará formado por los siguientes nodos:

- `SPEC`: será el nodo raíz del cual tendrá como hijo `typedef` o `declaration`.
- `typedef`: nodo del cual tendrá como primer hijo un nodo `type`, seguido de una serie de nodos `keyval` y como último hijo será un nodo `ID` que representará el nombre de la definición.
- `declaration`: tendrá como hijo dos nodos `ID` que el primero contendrá alguno de los tipos definidos anteriormente en `typedef` y el segundo representará el nombre de la variable en cuestión.

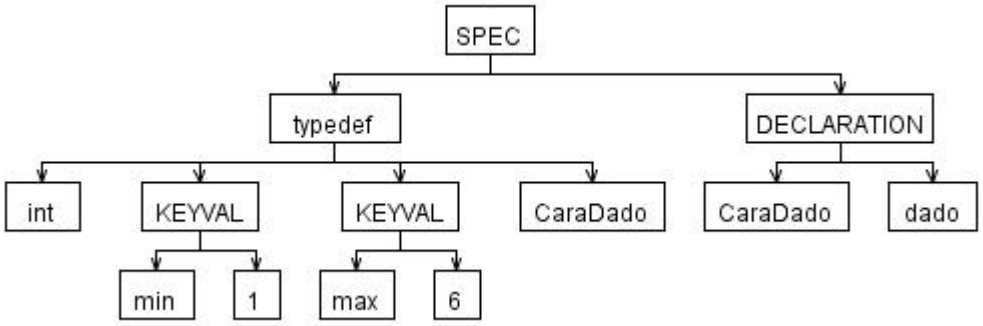

Figura 4.4.: Ejemplo de AST

- `type`: contendrá alguno de los tipos que acepta el sistema definido con anterioridad.
- `keyval`: nodo que tendrá como hijo dos nodos que representará el par clave-valor, donde clave representa el tipo de restricción y valor la asignación que tendrá.
- `ID`: nodo que representara un identificador de los anteriormente mencionados.

Para la implementación del analizador se ha utilizado la herramienta *ANTLRWorks* [13], que genera un analizador descendente $LL(*)$ a partir de una gramática en notación BNF. Para más información, se recomienda consultar A.4. En la figura 4.5 podemos observar el diagrama de clases de los analizadores.

4.5. Generador

En este bloque estructural, es el que da soporte a la generación de los datos de forma aleatoria, se podría decir que es la parte mas importante del programa. Toma las estructuras creadas en el parte y genera los datos con las restricciones deseadas.

Para crear dichos datos aleatorios, se ha utilizado el patrón de diseño software patrón Visitante y la técnica «Double-Dispatch» [14], en español doble despacho. A continuación veremos en que consiste.

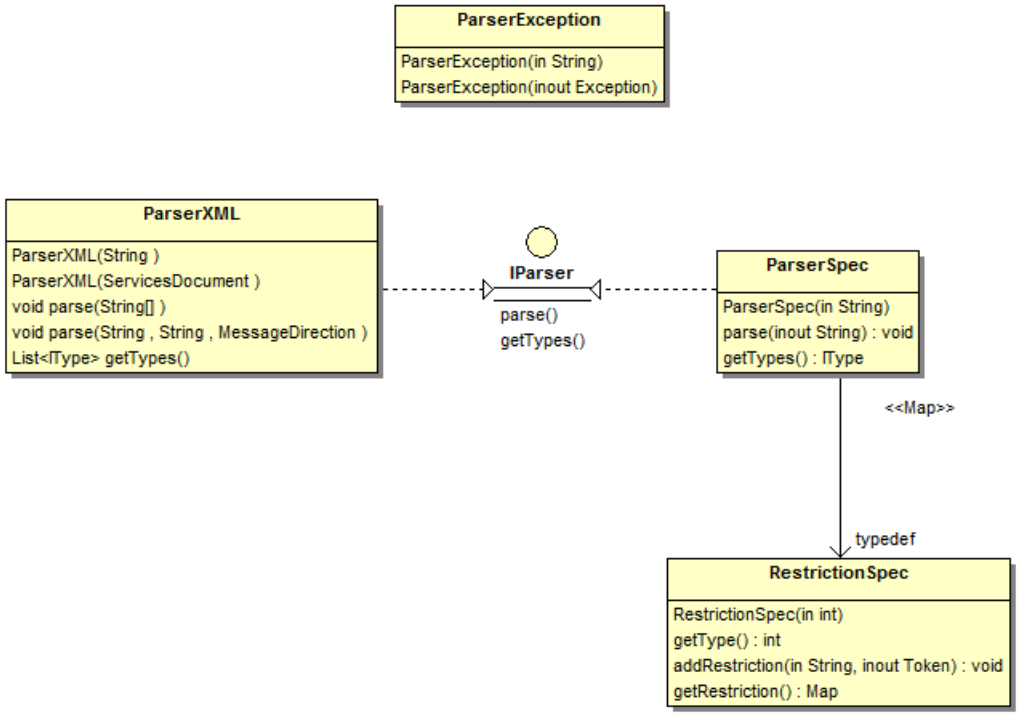

Figura 4.5.: Diagrama del analizador

4.5.1. Patrón Visitante

Se trata de un patrón de comportamiento, cuyo propósito es representar una operación sobre elementos de una estructura de objetos, permitiendo definir una nueva operación sin cambiar las clases de los elementos sobre los que opera. Estos son los elementos que intervienen en dicho patrón:

Visitante: interfaz que declara una operación *Visitar* para cada clase de operación *Elemento Concreto* de la estructura de objetos. El *Visitante* puede acceder al elemento directamente a través de su interfaz particular.

Visitante Concreto: implementa cada operación declarada por la interfaz *Visitante*. Proporciona el contexto para el algoritmo y almacena el estado local (normalmente acumula los resultados durante el recorrido de la estructura).

Elemento: define una operación *Aceptar* que toma un *Visitante* como argumento.

Elemento Concreto: implementa una operación *Aceptar* que toma un *Visitante* como argumento.

Estructura de objetos: puede ser una composición o colección de objetos (una lista o un conjunto, por ejemplo), en la que sus elementos se pueden enumerar. Proporciona una interfaz de alto nivel para que los visitantes visiten sus elementos.

Un cliente que usa el patrón Visitante debe crear un objeto *Visitante Concreto* y a continuación, recorrer la estructura, visitando cada objeto con el visitante. Cada vez que se visita un elemento, éste llama a la operación del *Visitante* que se corresponde con su clase. El elemento se pasa a sí mismo como argumento de la operación para permitir al *Visitante* acceder a su estado, si es necesario. Para recorrer la estructura, cada *Elemento Concreto* llama al método *Aceptar* de sus descendientes con el *Visitante Concreto*.

Conceptualmente, la estructura del código es la que aparece en el listado 4.4.

Listado 4.4: Estructura del código del patrón Visitante

```
1  class Node {
2  ...
3          void accept(Visitante v) {
4                  for each child of this node {
5                          child.accept(v);
6                  }
7                  v.visit(this);
8          }
9  }
10
11 class Visitor {
12 ...
13         void visit(Node n) {
14                 perform work on n
15         }
16 }
```

4.5.2. Características del patrón Visitante

1. Facilita la definición de nuevas operaciones: añadir una nueva operación conlleva únicamente añadir un nuevo VisitanteConcreto.

2. Agrupa operaciones relacionadas y separa las que no lo están: el comportamiento similar no está desperdigado por las clases que definen la estructura de objetos sino que está localizado en un Visitante. Las no relacionadas, se dividen en sus propias subclases del Visitante. Esto simplifica tanto las clases que definen los elementos, como los algoritmos definidos por los Visitantes. Cualquier estructura de datos específica de un algoritmo puede estar oculta en el Visitante.

3. Añadir nuevas clases elemento concreto es costoso: cada elemento concreto nuevo da lugar a una nueva operación abstracta del Visitante y a su correspondiente implementación en cada clase VisitanteConcreto.

4. Permite atravesar varias jerarquías de clases: a diferencia de un iterador, este patrón puede visitar objetos que no están relacionados por un padre común.

5. Permite acumular el estado: los visitantes pueden acumular estado a medida que van visitando cada elemento de la estructura de objetos, en vez de pasarlo como argumento o usar variables globales.

6. Rompe la encapsulación: el enfoque de este patrón arquitectónico asume que la interfaz de elemento concreto es lo suficientemente potente como para que los visitantes hagan su trabajo. Como consecuencia, el patrón suele obligarnos a proporcionar operaciones públicas que accedan al estado interno de un elemento, lo que puede comprometer su encapsulamiento.

4.5.3. Doble despacho

La clave del patrón visitante es el doble despacho («Double-Dispatch»). Esta es la técnica que permite añadir operaciones a las clases sin tener que modificarlas. Dada una operación, se dice que es de doble despacho si su ejecución depende de la clase de

petición y de los tipos de dos receptores. En el caso del patrón Visitante, la petición es el Aceptar y los receptores el visitante y el elemento. Un visitante debe recorrer cada elemento de la estructura de objetos. La responsabilidad del recorrido puede recaer sobre:

- La propia estructura de objetos: una colección iterará sobre sus elementos simplemente invocando a la operación Aceptar de cada uno de ellos. Esta es la opción más común.
- Un objeto iterador aparte: se puede hacer uso de los iteradores que ofrecen algunos lenguajes como C++.
- El Visitante: poner el algoritmo de recorrido en el visitante tiene la finalidad de implementar un recorrido especialmente complejo que dependa de los resultados de las operaciones de la estructura de objetos.

En nuestro proyecto, se ha utilizado la técnica del doble despacho («Double-Dispatch») para generar los datos aleatorios. En nuestro caso el elemento sería una de las estructuras de datos que guardan la información de los tipos y restricciones, con la petición que sería generar un dato aleatorio y el visitante sería la clase encargada de generar dichos datos de forma aleatoria.

En la figura 4.6 podemos ver el diagrama de clases del generador.

4.6. Formateadores

Es el encargado de formatear los datos generados en el bloque anterior a un formato especifico. Los formatos a elegir pueden ser CSV o Apache Velocity.

El formato CSV (del inglés «comma-separated values») representa los datos en forma de tabla donde las columnas se encuentran separadas por comas ’,’ y las filas por salto de línea. En nuestro caso, la primera tupla va a representar el nombre de las variables y cada tupla posterior representará un caso de prueba. Aunque este formato es muy cómodo de usar es relativamente sencillo, ya que no permite representar por ejemplo listas o

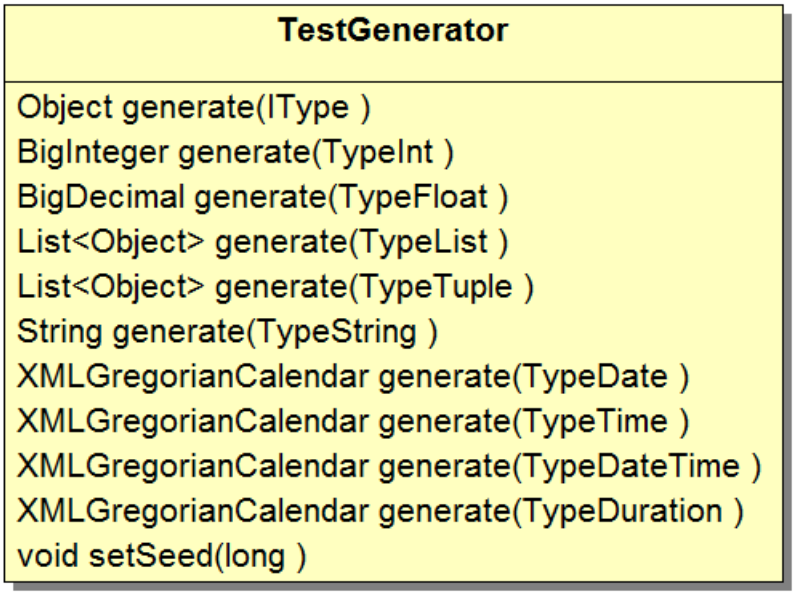

Figura 4.6.: Diagrama del generador

tuplas. Podemos ver un ejemplo de cómo quedaría una salida de nuestro programa en el listado 4.5.

Listado 4.5: Ejemplo de fichero de datos CSV generado por TestGenerator

```
1  as_reply, accepted, ap_limit, as_limit, ap_reply, req_amount
2  "silent", "true", 62309, 2234, "silent", 178305
3  "low", "true", 103284, 2084, "true", 176109
4  "high", "true", 85487, 2831, "true", 63107
5  "high", "false", 104922, 2719, "smart", 124955
6  "high", "true", 16135, 2943, "true", 150697
7  "smart", "false", 33981, 1788, "true", 40718
8  "smart", "true", 67753, 1345, "silent", 145627
9  "smart", "true", 110294, 3692, "false", 167525
10 "low", "true", 11569, 3661, "smart", 198891
11 "silent", "false", 110002, 2223, "false", 87784
```

El formato Apache Velocity es un motor de plantilla basado en Java, propone un for-

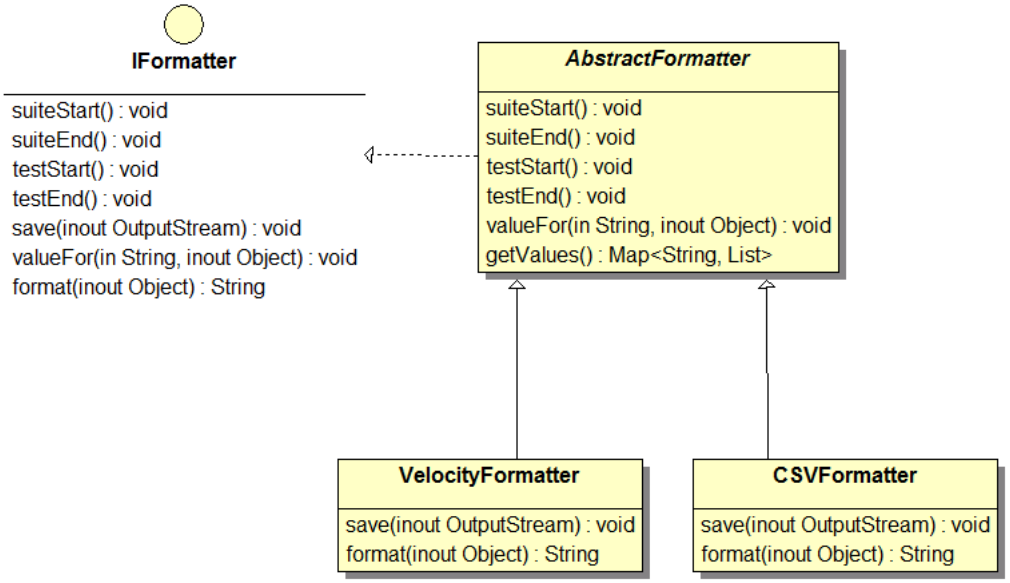

Figura 4.7.: Diagrama del formateador

mato donde se verá una variable y el conjunto de valores que va a tomar. Lo explicaremos desde el ejemplo

Listado 4.6: Ejemplo de fichero de datos Apache Velocity generado por TestGenerator

```
1 #set($as_reply = ["silent", "low", "silent", "low", "high"])
2 #set($accepted = ["true", "true", "true", "false", "false"])
3 #set($ap_limit = [91672, 88086, 90953, 111824, 41495])
4 #set($as_limit = [3496, 1371, 468, 3286, 526])
5 #set($ap_reply = ["true", "false", "true", "true", "false"])
6 #set($req_amount = [122740, 114966, 151674, 73523, 161326])
```

La estructura sigue el patrón `#set($NombreVariable = [valor1, valor2])`. Cada valor será un caso de prueba, si queremos incluir una lista como caso de prueba, sus elementos vendrán entre corchetes '[' ']' reservando la pareja de corchetes más general para especificar los valores.

En la figura 4.7 podemos observar el diagrama de clases de los formateadores.

Capítulo 5

Implementación y pruebas

En este punto hablaremos de los aspectos más técnicos, cuáles han sido las herramientas y tecnologías que se han elegido para el desarrollo de la solución. También hablaremos de las pruebas que se han llevado a cabo.

Este proyecto se ha utilizado el paradigma de programación orientado a objetos, en particular se ha utilizado el lenguaje Java. La elección del lenguaje viene como un requisito del grupo para poder reutilizar código existente (*ServiceAnalyzer*) como también pueda ser este código reutilizado en proyectos futuros.

Las grandes ventajas que nos aporta el paradigma de programación orientado a objeto [15] son las siguiente:

- Cercanía de sus conceptos a los del mundo real.
- Proceso de desarrollo más sencillo y más rápido.
- Facilita reutilización de diseño y códigos.
- Modificaciones, extensiones y adaptaciones más sencillas
- Sistemas más estables y robustos.

5.1. Integración continua

Como ya hemos mencionado anteriormente, el grupo UCASE trabaja con un modelo de integración continua.

El esquema básico montado en el grupo, consta de la interacción de cinco herramientas de distribución libre: *Jenkins*, *Subversion*, *Maven*, *Nexus* y *Sonar*.

5.1.1. Jenkins

Jenkins es una versión de la herramienta *Hudson* creada por la comunidad de software libre tras disputas con Oracle acerca del control de la marca registrada y la infraestructura del proyecto. *Hudson* es una herramienta de integración continua de código abierto, creada por Kôsuke Kawaguchi (antiguo empleado de Sun) en su tiempo libre, que se encarga de monitorear la ejecución de tareas repetidas, tales como la creación de un proyecto o la ejecución de tareas automáticas. Sus principales características son:

- Es fácil de instalar, ya que se distribuye como un fichero war.
- Todas las tareas de administración se realizan usando una interfaz web, lo cual facilita enormemente su configuración.
- Soporta notificación vía IM (Instant Messaging), e-mail y RSS (Really Simple Syndication).
- Genera gran cantidad de informes para JUnit y TestNG.
- La herramienta puede extenderse y personalizarse fácilmente mediante «plugins».
- Soporta CVS y Subversion para el control de cambios.

5.1.2. Subversion

Subversion es un sistema que permite almacenar todas las versiones de un árbol de ficheros, pudiendo así manipular todas las revisiones de cualquier fichero en cualquier

momento. Además de servir como una medida de seguridad contra la pérdida de información accidental, agiliza los cambios drásticos, ya que no hay que establecer medidas especiales por si fallaran: se pueden revertir los cambios hechos en cualquier momento.

Subversion es un sistema de control de versiones creado para reemplazar a *CVS*. Las ventajas que nos aporta en comparación con *CVS* son:

- Puede mantener revisiones de directorios completos.
- Establecer propiedades especiales sobre los elementos del repositorio.
- Enviar nuevas revisiones de forma atómica.

5.1.3. Maven 3

Maven es una herramienta para automatizar la gestión de software escrito en Java, incluyendo compilación, pruebas y despliegue, entre otras tareas. Fue creada dentro del proyecto Jakarta (2002), aunque actualmente el proyecto pertenece a la Apache Software Foundation. Posee una funcionalidad similar a la de *Apache Ant*. La diferencia principal entre ambas es que en *Ant* las acciones a realizar se definen en forma procedural paso por paso, mientras que con *Maven* se declara qué «plugins» se van a utilizar, con qué configuración y con qué dependencias y *Maven* se encarga del orden en el que se utilizan las cosas para lograr el objetivo declarado.

Maven utiliza un POM (Project Object Model) para describir el proyecto software a construir, sus dependencias con otros módulos y componentes externos, así como el orden de construcción de los elementos. Es un archivo basado en un formato XML que se ubica en la raíz del proyecto módulo. Viene con objetivos predefinidos para realizar ciertas tareas tales como la compilación del código y su empaquetado. Una característica clave de *Maven* es que está listo para usar en red. Nos permite publicar fácilmente binarios que incluyen todas las dependencias, de forma que sea simplemente descargar y utilizar. Si se desea utilizar una nueva biblioteca, por ejemplo, sólo tendrá que añadirse a las dependencias (al estilo de un paquete Debian) y dejar que *Maven* las descargue cuando haga falta.

El funcionamiento de *Maven* se basa en el uso de un repositorio a donde ir a buscar las dependencias. Cuando *Maven* sale a buscar y consigue una dependencia la guarda en el repositorio local que es un directorio en la máquina del usuario (normalmente *~/.m2/repository* en sistemas basados en UNIX). Las siguientes veces que necesite esta dependencia la obtendrá del directorio local, haciéndolo mucho más rápido que la primera vez.

Otra gran ventaja de *Maven* es que ayuda a imponer que los proyectos sigan una estructura uniforme. Las opciones de compilación, empaquetado, etc. pueden hacerse comunes a todos los proyectos, dejando una configuración mucho más consistente e independiente de los detalles de cada proyecto. *Maven* también puede generar proyectos para *Eclipse* y *NetBeans*, con lo que sólo tenemos que mantener un sistema.

Otro aspecto interesante de *Maven* es que, a partir de la información del POM y del código fuente, proporciona al usuario información de los proyectos que puede llegar a ser muy útil: árboles de dependencias, listas de direcciones, informes de las pruebas, referencias cruzadas entre las fuentes, etc. Asimismo, proporciona guías de buenas prácticas para el desarrollador.

5.1.4. Nexus

La distribución de *Maven* por defecto se descarga los artefactos del repositorio principal de *Maven*. Si bien esto a nivel personal es factible, cuando varios desarrolladores se tienen que descargar los mismos jars pesados una y otra vez es un despilfarro de ancho de banda y de tiempo considerable. Por otro lado, a una organización podría interesarle controlar o restringir de algún modo los artefactos que pueden descargarse los desarrolladores (para que descarguen una misma versión de una biblioteca, para que no usen el repositorio con fines ajenos a la organización, para que sólo empleen dependencias con una licencia compatible a la del proyecto en el que trabajan, etc.). Por ello se suele instalar un gestor repositorio propio, siendo *Nexus* una de las mejores opciones. Entre los motivos por los que debe elegirse Nexus, podemos destacar:

- Puesto que está creado por desarrolladores de *Maven*, la compatibilidad con esta

herramienta y la eficiencia en las comunicaciones con su repositorio central son las máximas posibles.

- Es pionero en el formato de repositorio índice.
- Su configuración y mantenimiento son sencillos. Sin embargo, puede ser usado de manera profesional y permite su extensión mediante «plugins».
- Destaca por poseer el modelo de seguridad más robusto y configurable de entre los gestores de repositorios actuales.
- *Nexus* usa Apache Lucene para indexar y buscar en tiempo real, sin necesidad de tener repositorios de contenido o bases de datos.

5.1.5. Sonar

Proyecto que se autodenomina como “una plataforma para la administración de la calidad del código”. Se trata de una herramienta que permite automatizar el análisis del código para gestionar aspectos tales como las nomenclaturas requeridas por una arquitectura o una metodología, el uso de buenas prácticas de programación, la repetición de código, el porcentaje de código cubierto por pruebas, ciertos parámetros de complejidad de clases y métodos, el porcentaje de código comentado, la evolución de las métricas a lo largo del tiempo, etc. En concreto, la plataforma *Sonar* permite gestionar la calidad del código controlando los siete ejes principales de dicha calidad del código:

- Arquitectura y diseño.
- Duplicaciones.
- Pruebas unitarias.
- Complejidad.
- Errores potenciales.
- Reglas de codificación.

- Comentarios.

5.2. Implementación de los analizadores

Como ya se ha mencionado anteriormente, este proyecto recibirá dos tipos de ficheros, los *wsdl* que mediante la herramienta *ServiceAnalyzer* se generan un catálogo el cual deberemos recorrer y por otro lado los ficheros del lenguaje diseñado especialmente para este proyecto, TestSpec.

5.2.1. Analizador del WSDL

Para el recorrido del catálogo generado a partir de los ficheros *wsdl*, se ha utilizado las clases de *ServiceAnalyzer* que utilizan la herramienta *XMLBeans*, dicha herramienta nos transforma un fichero con formato XML a una clase Java, la cual es mas fácil de trabajar y manipular.

Podemos ver un ejemplo de código de cómo recorrer dicho catálogo en el listado 5.1.

Los catálogos se estructuran en servicios, dentro de estos puede existir un conjunto de operaciones y dentro de estas podemos encontrar las variables de tipo de entrada («input»), de salida («output») o de error («fault»). El programa recibirá que servicio, operación y el tipo, y a partir de esto analizaremos la correspondiente sección de declaración donde podemos ver el tipo y restricciones de las variables que debemos generar. Esta sección es la que nos interesa para guardar la información en nuestra estructura interna de tipos para luego a partir de estos generar los datos aleatorios.

5.2.2. Analizador de los ficheros TestSpec

En un apartado anterior 4.4 pudimos observar cómo sería la forma de los ficheros *spec*. El formato de estos ficheros está descrito mediante una gramática. Veamos cuáles son las reglas que definen dicha gramática en el listado 5.2.

Una vez que tenemos dicha gramática, gracias a un «plugin» de *Maven*, podemos generar el código del analizador diseñado con la herramienta *ANTLRWorks*. Con dicho

Listado 5.1: Recorrido de un catálogo

```
String filePath = "messageCatalog.xml";
File inputXMLFile = new File(filePath);

ServicesDocument servicesDoc = ServicesDocument.Factory.parse(inputXMLFile);

Services services = servicesDoc.getServices();
TypeService[] servicesArray = services.getServiceArray();
for (int i = 0; i < servicesArray.length; ++i) {
       System.out.println("Service " + servicesArray[i].getName());
       TypePort[] portsArray = servicesArray[i].getPortArray();
   for (int j = 0; j < portsArray.length; ++j) {
       System.out.println("\tPort " + portsArray[j].getName());
       TypeOperation[] operationsArray = portsArray[j].getOperationArray();
       for (int k = 0; k < operationsArray.length; ++k) {
           System.out.println("\t\tOperation " + operationsArray[k].getName());
           TypeInput input = operationsArray[k].getInput();
           TypeVariable[] variables = input.getDecls().getVariableArray();
           System.out.println("\t\t\tInput declarations variable #1"
                                                  + variables[0].getType());
           System.out.println("\t\t\tInput template " + input.getTemplate());
       }
   }
}
```

código podemos recorrer los ficheros y crear nuestras estructuras de datos que guardan las restricciones de los datos que posteriormente vamos a generar de forma aleatoria.

Listado 5.2: Gramática del lenguaje TestSpec

```
1  spec: (line ';')*
2         ;
3
4  line: typedef
5         | declaration
6         ;
7
8  typedef: TYPEDEF type restrictions? ID;
9  declaration: type ID;
10
11 type
12        : TYPE_INT
13        | TYPE_TUPLE
14        | TYPE_LIST
15        | TYPE_FLOAT
16        | TYPE_STRING
17        | TYPE_DATE
18        | TYPE_DATETIME
19        | TYPE_TIME
20        | TYPE_DURATION
21        | ID
22        ;
23
24 restrictions: '(' (restriction (',' restriction)* )? ')';
25
26 restriction: ID '=' restriction_value
27
28 restriction_value
29        : INT
```

```
30        | STRING
31        | FLOAT
32        | type
33        | '{' (restriction_value (',' restriction_value)* )? '}'
34        ;
35
36  TYPEDEF        : 'typedef';
37  TYPE_INT       : 'int';
38  TYPE_TUPLE     : 'tuple';
39  TYPE_LIST      : 'list';
40  TYPE_FLOAT     : 'float';
41  TYPE_STRING    : 'string';
42  TYPE_DATE      : 'date';
43  TYPE_DATETIME  : 'datetime';
44  TYPE_TIME      : 'time';
45  TYPE_DURATION  : 'duration';
46
47  ID  :   ('a'..'z'|'A'..'Z'|'_') ('a'..'z'|'A'..'Z'|'0'..'9'|'_')*
48      ;
49
50  INT :   ('+'|'-')?'0'..'9'+
51      ;
52
53  FLOAT
54      :   ('+'|'-')?('0'..'9')+ '.' ('0'..'9')* EXPONENT?
55      |   '.' ('0'..'9')+ EXPONENT?
56      |   ('+'|'-')?('0'..'9')+ EXPONENT
57      ;
58
59  EXPONENT : ('e'|'E') ('+'|'-')? ('0'..'9')+ ;
60
61  COMMENT
62      :   '//' ~('\n'|'\r')* '\r'? '\n'
```

```
63      |   '/*' ( options {greedy=false;} : . )* '*/'
64      ;
65
66  STRING
67      :  '"' ( ESC_SEQ | ~('\\'|'"') )* '"'
68      ;
```

5.3. Generadores de datos aleatorios

En esta apartado, veremos cómo hemos generado los datos de forma aleatoria. Para poder generar dichos datos, nos hemos basado en la clase *Random* del paquete *java.util* y se encuentra disponible desde la versión JDK 1.0.

La clase *Random* se utiliza para generar secuencias de números pseudoaleatorios. Dicha clase utiliza una semilla de 48 bits que es modificada mediante una fórmula de congruencia lineal. Si dos instancias de la clase *Random* tiene la misma semilla y se llama a sus métodos de la misma forma, generarán los mismos números.

5.3.1. Enteros

De forma inmediata, podemos pensar que la mejor forma de representar tipos enteros, debería ser generando de forma aleatoria instancias del tipo int de Java. En realidad así fue como se hizo en un primer lugar hasta que viendo los ficheros *wsdl* que tiene el grupo, pude observar que quería representar el tipo unsigned int, este tipo toma valores en el intervalo [0, 4.294.967.295] este rango no concuerda con el tipo int de Java que sus valores van en el intervalo [-2.147.483.648, 2.147.483.647]. Como solución a dicho problema decidimos usar la clase *BigInteger* localizada en el paquete *java.math*, se trata de una clase que representa enteros de precisión arbitraria.

Para generar un *BigInteger* aleatorio, podemos usar unos de sus constructores que los genera de forma aleatoria, pero esto no es tan sencillo puesto que el constructor recibe el número de bit que necesita para representar dicho número. Con esto hay que tener

cuidado, ya que podemos salirnos del rango en el cual lo necesitamos. Para generarlo hacemos lo siguiente:

1. Calculamos la diferencia del intervalo, valor máximo permitido menos el valor mínimo permitido.
2. Generamos el número de forma aleatoria con el número de bits de la diferencia (máximo-mínimo).
3. Sumamos el valor mínimo al número aleatorio resultante.
4. Comprobamos que el número no es mayor que máximo. Si fuese así, volvemos al paso 2.

Veámoslo mejor con un ejemplo. Supongamos que tenemos que generar un número aleatorio en el rango [20, 50]: pasaremos a ejecutar los pasos anteriormente descritos. Calculamos la diferencia del intervalo y nos obtenemos como resultado el número 30 (50- 20 = 30).

Para representar el número 30 son necesarios 5 bits ($2^5 = 32$), por lo que generamos números de forma aleatoria con 5 bits, lo que nos da la posibilidad de representar 32 números. Supongamos que de forma aleatoria se genera el número 13, pasaríamos a sumárselo a la cota inferior del intervalo obteniendo el número 33 (20 + 13 = 33) por lo que este sería nuestro número generado aleatoriamente.

Si por el contrario hubiésemos generado de forma aleatoria el número 31, al sumárselo a la cota inferior del intervalo obtendríamos el número 51 (20 + 31 = 51). Dicho número supera la cota superior del intervalo por lo que tendríamos que generar otro número aleatoria hasta que no superemos el intervalo.

5.3.2. Números en coma flotante

De la misma forma que podemos pensar que para representar un entero podíamos usar el tipo int de Java, podemos pensar que para representar los números de coma flotante, podemos usar el tipo float de Java. Esto no es así ya que esto enmascara un

problema un tanto difícil de encontrar y es la forma de representarlo. Los tipos float y double de representan los números en coma flotante en base 2 y no en base 10, esto es así por cuestiones de eficiencia ya que el procesador le es más fácil trabajar con este tipo de representación, Java usa el estándar IEEE 754 para representar dichos números. Esto hace que algunos números no se puedan representar de forma exacta, por ejemplo, el número 0.1 en base 10 es exacto, pero periódico en base 2. Es como intentar representar 1/3 en base 10 resulta un número periódico (0'333333...).

Para resolver este problema, debemos usar un tipo que represente los números en formato decimal: *BigDecimal*, del paquete *java.math*. Este tipo está recomendado por ejemplo para operaciones que impliquen suma de dinero. El problema es que como la representación de cada sumando es inexacta, el resultado acumula los errores de cada uno de esos sumandos. Eso quiere decir que al final, debido al error acumulado, el resultado puede no tener las cotas de exactitud que necesitamos. Internamente *BigDecimal* es un *BigInteger* con una escala determinada que indica la posición del separador decimal.

Para calcular un número en coma flotante de forma aleatoria usamos la fórmula `(max - min) * rnd.nextFloat() + min`, siendo `max` el valor máximo permitido, `min` el valor mínimo permitido y `rnd.nextFloat()` un número calculado de forma aleatoria entre $0{,}0$ y $1{,}0$.

5.3.3. Cadenas

Para generar cadenas de caracteres de forma aleatoria hemos usado la clase *string* de Java para su representación. Podemos generarlo usando una expresión regular. Para ello, nos hemos servido de *Xeger*.

Xeger es un programa que es capaz de generar una instancia de forma aleatoria a partir de una expresión regular. Dicho programa se encuentra bajo la licencia Apache License 2.0 al igual que este proyecto, con lo cual no tenemos ningún problema en utilizarlo.

Xeger es capaz de generar una cadena alfanumérica de forma aleatoria usando un autómata finito no determinista. Dicho autómata representa la expresión regular y en vez de dada una entrada recorrer el autómata para ver si llega a un estado final y validarlo,

lo que hace es que va pasando de estado a estado de forma aleatoria hasta llegar a un estado de aceptación y devuelve la cadena que ha generado.

Para generar una cadena alfanumérica aleatoria basta con pasarle la expresión regular `[A-Za-Z0-9]{0,longitud}`, donde `longitud` representa la máxima longitud que pueda tomar la cadena.

En la figura 5.1 podemos observar un autómata finito para explicar mejor el funcionamiento de *Xeger* con un ejemplo. Este autómata representa la expresión regular `a[aeo]*`. *Xeger* actuaria de la siguiente forma: entraría por el estado inicial y pasaría al estado q_1, ahora de forma aleatoria decidiría si realizar una transición al estado q_2, q_3 o al propio q_1 o por el contrario, tomarlo como salida válida. De la misma forma seguiría actuando hasta que de forma aleatoria decida tomar un estado final.

5.3.4. Fechas

Para representar las fechas, hemos utilizado el tipo Java *XMLGregorianCalendar*. Este tipo es el más interesante para representar nuestras fechas, puesto que se corresponde con la sintaxis que utiliza XMLSchema, aunque para poder generar fechas de forma aleatoria hemos utilizado la clase *Calendar*.

El proceso consiste en tomar nuestras fechas máximas y mínimas representadas por instancias de la clase *XMLGregorianCalendar* y pasarla a la clase *Calendar*, esto es así porque la clase *Calendar* tiene un método que nos permite representar la fecha en milésimas de segundo con el método *getTimeInMillis*, con esto lo que hacemos es generar un número aleatorio comprendido entre los milésimas de segundo de la fecha mínima y la fecha máxima y a continuación volver a convertir ese número generado aleatoriamente al tipo *Calendar* y este a su vez al tipo *XMLGregorianCalendar*.

5.3.5. Contenedores

Para generar las clases contenedoras de forma aleatoria, basta con generar el número de objetos que pueda albergar de forma aleatoria del tipo correspondiente.

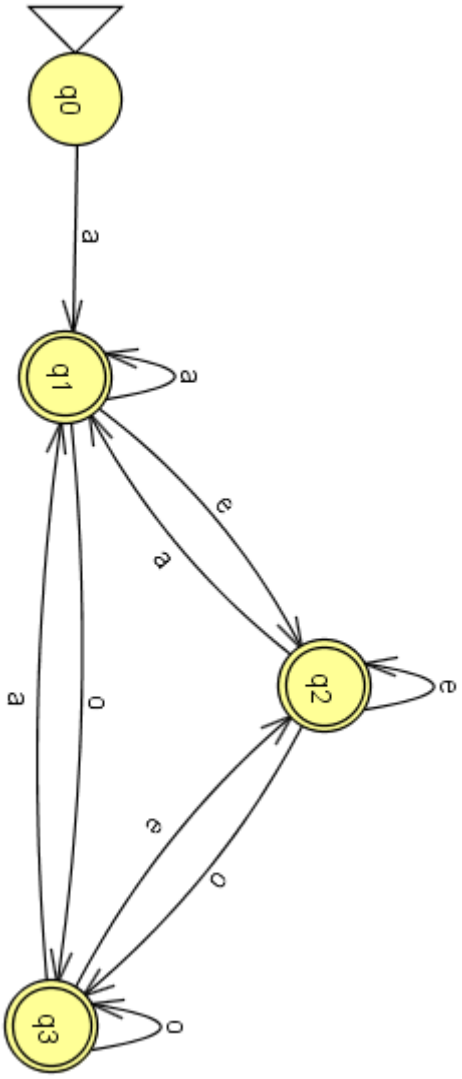

Figura 5.1.: Ejemplo de automata finito generado por Xeger para la expresión regular a[aeo]*

5.4. Pruebas

Una de las partes fundamentales en cualquier aplicación, es el proceso de pruebas. Esto permite aumentar la calidad del sistema reduciendo el número de errores. También permite disminuir la probabilidad de que el sistema falle después de realizar alguna modificación. Existen distintos tipos de pruebas:

- Pruebas unitarias: son aquellas pruebas diseñadas para probar partes muy específicas de la aplicación. Comprueban de forma automática de un conjunto reducido y cohesivo de clases. Estas pruebas son diseñadas por los programadores.
- Pruebas de aceptación: están destinadas a comprobar la funcionalidad del sistema, es decir verifican si cumplen las expectativas pedidas por el cliente.
- Pruebas de integración: son aquellas pruebas que verifican que el sistema funciona de la forma esperada dentro del marco del sistema. En este proyecto, estas pruebas tienen un grado alto de importancia, ya que dicho programa se utilizará dentro del grupo UCASE y deberá funcionar de la forma esperada aunque cambie algún programa del cual depende este proyecto.
- Pruebas de implantación: desde el inicio de un proyecto, el sistema debe de implantarse en un entorno similar al real, posiblemente a menor escala, y comprobarse su correcto funcionamiento. Este tipo de prueba es muy fácil de seguir en el grupo UCASE debido a su sistema de integración continua.

5.4.1. Naturaleza de las pruebas

La totalidad de las pruebas usadas son pruebas de caja negra, por ser fáciles de mantener a pesar de cambios en la implementación.

Recordemos que las pruebas de caja negra son aquellas que se centran en lo que se espera de un módulo, es decir, intentan encontrar casos en los que el módulo no se atiene a su especificación. Para ello, se apoyan en la especificación de requisitos del módulo, por lo que también se denominan pruebas funcionales. Conociendo una función específica

para la que fue diseñado el producto, se pueden diseñar pruebas que demuestren que cada función está bien resuelta. El probador se limita a suministrarle datos como entrada y estudiar la salida, sin preocuparse de lo que pueda estar haciendo el módulo por dentro.

Las pruebas de caja negra son útiles en cualquier módulo del sistema pero están especialmente indicadas en aquellos que van a servir de interfaz con el usuario.

El problema con las pruebas de caja negra no suele estar en el número de funciones proporcionadas por el módulo (que siempre es un número muy limitado en diseños razonables), sino en los datos que se le pasan a estas funciones. El conjunto de datos posibles suele ser muy amplio.

A la vista de los requisitos de un módulo, se sigue una técnica algebraica conocida como “clases de equivalencia”. Esta técnica trata cada parámetro como un modelo algebraico donde unos datos son equivalentes a otros. Si logramos transformar un rango excesivamente amplio de posibles valores reales en un conjunto reducido de clases de equivalencia, entonces es suficiente probar un caso de cada clase, pues los demás datos de la misma clase son equivalentes. Es importante identificar qué rangos de datos pueden alterar el comportamiento del programa y así definir zonas de trabajo. Es imprescindible pasar pruebas con al menos un dato de cada zona, tanto si el programa debe funcionar como si debe dar un mensaje de error. La experiencia indica, además, que suelen producirse fallos en los bordes de las zonas, por lo que se recomienda probar siempre con datos extremos.

Esta técnica es en la que nos hemos basado a la hora de elaborar las pruebas del sistema.

Los casos de prueba se han definido utilizando el framework *JUnit*, el conjunto de bibliotecas creadas por Erich Gamma y Kent Beck para hacer pruebas unitarias de aplicaciones Java.

5.4.2. Diseño de las pruebas

Dada a la forma que está estructurada la aplicación, podemos englobar las pruebas unitarias a tres elementos a probar:

- Analizadores: en este punto hay que comprobar que el programa crea las estructuras de datos adecuadas, guardando las restricciones que le corresponde a cada dato. Esto hay que comprobarlo por dos partes, una para los ficheros *wsdl* y otra para los ficheros *spec*.
- Generador: en este parte del programa hay que comprobar que los datos que se generan aleatoriamente cumplen con las restricciones impuestas por la estructuras de datos que almacenan la información, comprobar esto es un tanto difícil puesto que los datos se generan de forma aleatoria así que hay que intentar acotarlos en puntos críticos y jugar con generarlos un número de veces apropiada a cada situación.
- Formateadores: aquí comprobaremos si el formateo de la salida es el adecuado según el dato aleatorio generado. Hay que comprobarlo tanto para el formato CSV como para el de Apache Velocity.

Además de diseñar estas pruebas unitarias, es necesario diseñar pruebas de integración, con ella se pretende cubrir su correcta integración con *ServiceAnalyzer*, para ello hemos creado pruebas con los siguientes *wsdl* que se usan actualmente en el grupo:

- LoanApprovalRPC
- LoanStructured
- MetaSearch
- SquaresSum
- TacService
- LoanApprovalExtended

- MarketPlace
- ShippingServiceAsynchronous

Aparte de estas pruebas, también se han diseñado dos pruebas más de integración, con estas pruebas no se comprueba el resultado sino que se comprueba que el programa sigue su recorrido normal y no lanza ningún tipo de excepción al ejecutarse. Una de las prueba utiliza un fichero wsdl y el formato CSV y la otra utiliza un fichero *spec* y el formato Apache Velocity.

5.5. Validación

En esta fase se ha mejorado la calidad del código. Para mejorar la calidad del código se ha utilizado la herramienta *Sonar*. Dicha herramienta se encuentra alojada en los servidores que usa el grupo UCASE. Al subirse al repositorio una versión nueva del código, *Sonar* comprueba la calidad del mismo.

Sonar es capaz de detectar puntos débiles de nuestro proyecto como errores potenciales en el código, escasez de comentarios, clases demasiado complejas, escasez de cobertura de las pruebas unitarias, etc.

Lo primero que llama la atención es la sección de "Violations" que nos indica los errores que tiene nuestro código dividido en niveles de gravedad. Esta es una visión muy útil para asegurar que nuestro código está escrito de acuerdo a las buenas prácticas de Java mejorando así en eficiencia, usabilidad y mantenibilidad, fundamentalmente.

Sonar también da información del resultado de los test y de su cobertura; así como del porcentaje de líneas que son comentarios y de líneas duplicadas en el código. Este último dato nos puede servir para darnos cuenta de las zonas de la aplicación que están repetidas y que convendría refactorizar en una única clase.

Entre las correcciones realizadas a partir de las indicaciones Sonar podemos citar las siguientes:

- Uso de *StringBuilder*: esta clase se ha empleado en el generador de plantillas para optimizar las concatenaciones. Un *String* es un objeto inmutable, por lo que con

cada concatenación estamos creando un objeto nuevo. Si bien en pocos objetos no es importante, a la hora de trabajar con muchos *String*, por ejemplo, cuando la concatenación se produce dentro de un bucle, supondría un despilfarro de memoria. Java provee soporte especial para la concatenación de *String*s con la clase *StringBuilder*. Un objeto *StringBuilder* es una secuencia de caracteres mutable: su contenido y capacidad puede cambiar en cualquier momento.

- Complejidad ciclomática: con esto podemos detectar que métodos son demasiados complejos, si son demasiado complejos probablemente serán más difíciles de mantener, esto sugiere realizar una refactorización del mismo en la manera de lo posible siempre que el código no pierda su claridad.
- Número mágico: al usar ciertos números, *Sonar* interpreta que pueden tener un significado especial por lo cual sugiere llevarlo al uso de una constante que lo representa por si en un futuro dicho valor cambia.
- Corte de relaciones cíclicas: había dos paquetes con dependencias cíclicas, por lo que se movieron las clases conflictivas de un paquete al otro.

Capítulo 6

Conclusiones

6.1. Valoración personal

La elaboración de este proyecto ha supuesto una gran aportación a mis aptitudes y actitudes como ingeniero. Con este proyecto he realizado una primera toma de contacto con la clase de trabajo que tendré que afrontar en una futura vida laboral.

Una de los valores añadidos a mi persona que ha tenido desarrollar este proyecto, ha sido la colaboración con un grupo de investigación. Dicha experiencia ha sido muy enriquecedora puesto que no sólo he aprendido el uso de ciertas tecnologías, además de ello he podido comprobar cómo se trabaja en grupo asistiendo a distintos seminarios durante el desarrollo del proyecto. Esto ha facilitado aumentar mis competencias transversales como son el trabajo en equipo, expresión oral y escrita así como intervenciones en público.

Por primera vez he trabajado con un entorno de integración continua, valorando su utilidad y la importancia de poder acceder a la versión más reciente de código, sobre todo cuando existe dependencias entre distintos proyectos dentro del mismo grupo. De esta forma he conocido y aprendido a utilizar la herramienta *Maven* para gestión de proyectos, valorar y utilizar las herramientas de control de versiones como puede ser *Subversion*, a valorar la importancia de tener un código de calidad gracias a la herramienta *Sonar*, etc.

Todo esto también me ha servido para valorarme a mí mismo como ingeniero, ya que he estado utilizando herramientas las cuales eran en un principio totalmente desconocidas a mi persona por lo que he tenido que usar distintos manuales, artículos y otras fuentes para poder aprender a usar dichas herramientas. En cierta manera me ha ayudado a perderle un poco el miedo al mundo laboral por el desconocimiento que tengo del mismo.

En cuanto al lenguaje de programación utilizado, en este caso Java que para mí era totalmente desconocido, me ha servido para ver que en la universidad no me enseñaron C++, si no a programar y con un poco de esfuerzo me veo capaz de programar en el lenguaje que mejor satisfaga las características exigidas en cada proyecto

A través del proyecto, he aprendido a valorar la importancia que tiene realizar una buena batería de pruebas para obtener códigos más fiables y seguros gracias al uso de *JUnit*.

Cabe mencionar también el grado de comprensión que gracias a este proyecto he logrado alcanzar referente a los traductores y compiladores del lenguaje, puesto que en el mismo he tenido que desarrollar una gramática para posteriormente tener que realizar su análisis léxico y sintáctico con la herramienta *ANTLRWorks*, esto ha sido uno de los mayores retos de aprendizaje en mi proyecto.

LaTeX ha sido un elemento muy importante en la elaboración de la documentación del presente proyecto. Ha sido necesario aprenderlo partiendo de ningún conocimiento previo, gracias a ello he visto su utilidad y el porqué de su uso frente a los tradicionales procesadores de textos.

También decir que se ha realizado una importante labor de documentación (javadocs, manuales) y que se ha decidido que el proyecto sea Software Libre para que pueda resultar de provecho para la comunidad de desarrolladores.

Por último, dar las gracias al grupo de investigación UCASE y a todos sus componente (alumnos y profesores) puesto que me han ayudado en mi aprendizaje como ingeniero.

6.2. Trabajo futuro

Este proyecto va a tener continuidad dentro del marco de trabajo del grupo UCASE, siendo útil incluso para el desarrollo de nuevos proyectos. A corto plazo, se elaborarán especificaciones TestSpec para los casos de estudio utilizados normalmente por el grupo. A medio plazo, se empleará el lenguaje TestSpec para generadores basados en estrategias más avanzadas (por ejemplo, algoritmos evolutivos).

Se planea continuar este proyecto en el Segundo Ciclo en Ingeniería Informática, añadiendo diversas funcionalidades. En principio, se piensa dotarlo de una interfaz gráfica para que sea más cómodo de usar, incluir nuevos tipos de datos como pueden ser árboles, generar texto a partir de un diccionario predefinido de palabras y usar otras distribuciones estadísticas, entre otras ideas. Las posibilidades son varias y se abren varios hilos referentes a este proyecto.

Apéndice A

Generación de analizadores LL(*) con ANTLR

A.1. Introducción

En este manual hablaremos de qué son las gramáticas formales así como definiremos qué es un AST («Abstract Syntax Trees», o Árboles de Sintaxis Abstracta), su utilidad y cómo *ANTLR* puede ayudarnos a escribir nuestros AST.

A.1.1. Gramáticas formales

La gramática de un lenguaje es el conjunto de reglas capaces de generar todas las cadenas que pertenecen al lenguaje, ya sea éste un lenguaje formal o un lenguaje natural.

Una gramática formal se define como un conjunto formado por cuatro componentes [16]:

$$G = \langle T, N, S, P \rangle$$

- Un conjunto de «tokens», T, también llamados símbolos terminales, que son los símbolos básicos con los que se construye las cadenas. Los «tokens» constituyen los símbolos del alfabeto del lenguaje descrito por la gramática.

Listado A.1: Ejemplo de una gramática

```
1  instrucciones : (expresion ';')*
2                ;
3
4  expresion : exp_mult (('+'|'-') exp_mult)*
5            ;
6
7  exp_mult : exp_base (('*'|'/') exp_base)*
8           ;
9
10 exp_base : NUMERO
11          | '(' expresion ')'
12          ;
13
14 NUMERO: ('0'..'9')+
15       ;
```

- Un conjunto de símbolos no terminales, N. Es decir, variables sintácticas que designan cadenas de símbolos terminales.
- En toda gramática existe un símbolo no terminal destacado, S perteneciente a N, al que se le reconoce como el axioma, símbolo de comienzo o símbolo inicial. Al conjunto de cadenas que representa este símbolo se le conoce como el lenguaje generado por esta gramática.
- Un conjunto de producciones, P. Cada producción consta de una cadena de símbolos terminales (llamado el lado izquierdo o cabecera de la producción), un separador y otra secuencia de símbolos terminales y no terminales (llamado el lado derecho o cuerpo de la producción). Estas producciones describen la forma en que se pueden combinar los símbolos terminales y no terminales para formar cadenas.

Es muy importante desarrollar la capacidad de describir entradas mediante gramáticas

y de prever el lenguaje que genera una gramática dada. Esta habilidad sólo se desarrolla mediante la práctica viendo ejemplos y realizando ejercicios.

En el listado A.1 podemos observar una gramática en notación de Backus-Naur que es capaz de representar las operaciones matemáticas de suma, resta, multiplicación y división, además de representar asociatividad con el uso de los paréntesis.

Veamos sobre el ejemplo cada una de las partes de la gramática:

- El conjunto de «tokens» o símbolos terminales usados en esta gramática son los símbolos matemáticos '+', '-', '*', '/', '(', ')' y el símbolo terminal *NUMERO*. Como norma los símbolos terminales se suelen escribir en mayúsculas y los no terminales en minúsculas.

- El símbolo de comienzo de esta gramática es *instrucciones*.

- El conjunto de producciones está formado por *instrucciones*, *expresion*, *exp_mult* y *exp_base*.

Esta gramática acepta expresiones de la forma `(2+4)/3;`, respetando las asociatividades y prioridades de operadores habituales.

A.1.2. ANTLR

ANTLR es una herramienta capaz de generar analizadores sintácticos recursivos descendentes basados en gramáticas LL(*). Dicha herramienta está escrita íntegramente en Java y es capaz de generar Árboles de Sintaxis Abstracta en código Java o C++.

Podemos considerar los AST [16] como una representación resumida de un árbol de análisis que es útil para representar las construcciones del lenguaje. En un árbol de análisis se muestran todas las producciones reales utilizadas por nuestra gramática (sintaxis concreta). Es posible que muchas de esas producciones hayan surgido como consecuencia de la adaptación de la gramática al método de análisis, y son detalles que no interesa tener en cuenta. Para ello construimos una representación resumida de dicho árbol de análisis en la que ocultamos todos los detalles debidos al método de análisis

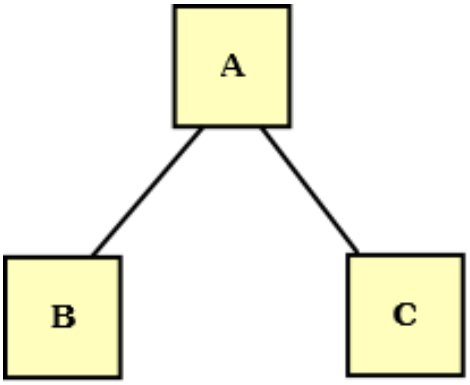

Figura A.1.: Arbol de la notación #(A B C)

sintáctico utilizado y sólo mostramos los detalles que son significativos en relación con la semántica de la sentencia.

Los AST pueden intervenir en varias fases del análisis: como producto del análisis sintáctico, como elemento intermedio en sucesivos análisis semánticos y como entrada para la generación de código.

La forma normal de representar un árbol en ANTLR utiliza una notación en la cual el árbol se representa mediante el uso de paréntesis, el primer elemento del paréntesis es el nodo padre y los sucesores nodos hijos. En la figura A.1 se muestra el AST representado por `#(A B C)`. Se trata de un árbol con "A" en la raíz, y los hijos B y C.

Esta notación puede anidarse para describir árboles de estructuras más complejas. En la figura A.2 podemos ver el árbol representado por `#(A B #(C D E))`. Tiene "A" en la raíz, "B" y "C" en el segundo nivel y "D" y "E" como hijos de "C".

En este anexo veremos los mecanismos que proporciona ANTLR para crear árboles de sintaxis abstracta. Estos aspectos (junto con la notación utilizada para recorrer dichos árboles) constituyen la parte más original de la herramienta. La solución aportada para describir dichos árboles a partir de la especificación sintáctica es realmente imaginativa y da lugar a descripciones bastante claras y compactas.

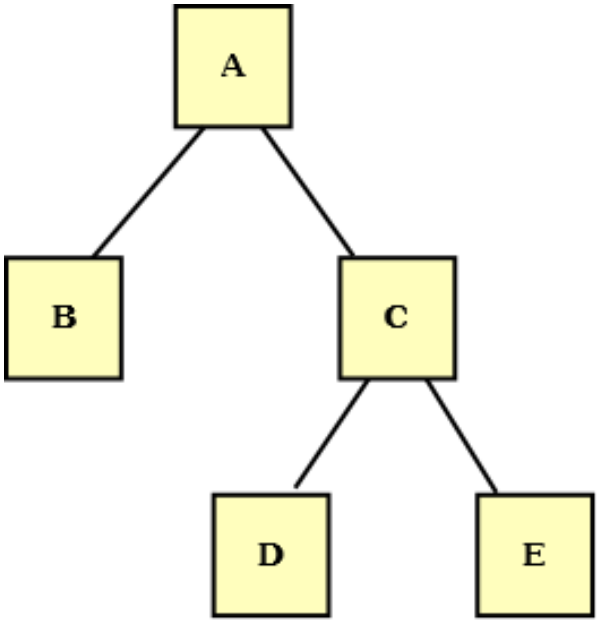

Figura A.2.: Árbol de la notación #(A B #(C D E))

A.2. Construcción del AST

ANTLR no requiere de una especificación adicional para definir los árboles de sintaxis abstracta. En su lugar inserta una serie de símbolos en la propia gramática concreta que establecen qué elementos merece la pena mantener en árbol de sintaxis abstracta y a partir de ahí lo genera de forma automática. De esta forma con tan sólo enriquecer la especificación del análisis sintáctico se tiene resuelta la especificación y construcción del árbol de sintaxis abstracta. En el listado A.2 veremos un ejemplo de especificación de un AST.

A simple vista la especificación anterior puede parecer un poco críptica, pero una vez que se comprende el significado de los distintos elementos en juego se descubre que resulta muy cómodo especificar la construcción de árboles con ellos:

- La sección `options` incluye la instrucción *output=AST*. De esta forma se activa la construcción automática del árbol de sintaxis abstracta.
- La sección `tokens` permite añadir nuevos «tokens» a los ya definidos en el análisis

Listado A.2: Ejemplo de un analizador sintáctico

```
1  grammar gramaticaprueba;
2  options
3  {
4    output=AST;
5  }
6  tokens
7  {
8    CUENTA;
9    SUMARESTA;
10   MULTDIV;
11 }
12 instrucciones : (expresion ';')* -> ^(CUENTA expresion*)
13                      ;
14
15 expresion : exp_mult ('+'|'-' exp_mult)* -> exp_mult ^(SUMARESTA exp_mult*)
16                 ;
17
18 exp_mult : exp_base ('*'|'/' exp_base)* -> exp_base ^(MULTDIV exp_base*)
19                ;
20
21 exp_base : NUMERO
22                | '('! expresion ')'!
23                ;
24
25 NUMERO
26     :   ('0'..'9')+
27     ;
```

léxico. En este caso serán símbolos terminales virtuales que servirán como raíz a árboles para los que en la gramática no hay ningún símbolo terminal que pueda desempeñar esta función. Esto suele ser muy común en las listas, en este caso es utilizado para generar un nodo padre con mayor sentido semántico.

- Cuando un símbolo va seguido del operador *!* se considerará no relevante y no se incluirá en el árbol de sintaxis abstracta.
- El operador `^` indica que el símbolo al cual acompaña debe ser la raíz del árbol. Si en una regla no se destaca ningún símbolo como raíz, el árbol generado será una secuencia de hermanos sin padre.
- Se puede deshabilitar la construcción automática y sustituirla por otra. Esto es lo que se hace en la regla de *instrucciones* para dotar de un padre a la secuencia de *expresion*.
- Con el elemento *^(CUENTA expresion*)* se crea un árbol cuya raíz es el símbolo terminal ficticio `CUENTA`, el cual tiene como hijos una series de nodos *expresion*.
- También podemos dejar la gramática sin tocar y especificar el árbol de sintaxis abstracta mediante el operador "->", como sucede en la regla *instrucciones*.

Para la expresión `5+4+6*2` generaría el AST A.3 donde:

- *NIL*: representa el inicio de evaluación de la expresión.
- *CUENTA*: es el símbolo inicial de la gramática.
- +: representa la operación suma.
- *: representa la multiplicación.

A.3. Manipulación del AST

El comportamiento de los árboles de sintaxis abstracta se establece a través de la interfaz *AST*. Por defecto los árboles son objetos de la clase *CommonAST* que implementa

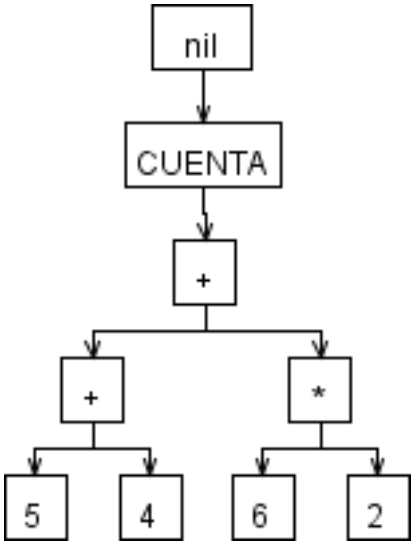

Figura A.3.: AST de la expresión 5+4+6*2

dicha interfaz, y que a su vez es subclase de *BaseAST*. La clase *BaseAST* implementa todos los métodos de la interfaz AST salvo los relacionados con la información contenida en las raíces que son implementados en *CommonAST* (permitiendo extender esta información con facilidad).

Algunos de los métodos de que disponen los árboles de sintaxis abstracta son:

- `AST getFirstChild()`: devuelve el primer hijo de un nodo interno de un árbol.
- `AST getChild(int n)`: devuelve el hijo `n` de un nodo interno del árbol.
- `AST getNextSibling()`: devuelve el hermano derecho de un árbol.
- `boolean equals(AST t)`: comprueba si las raíces de ambos árboles son iguales.
- `boolean equalsTree(AST t)`: comprueba si los árboles son iguales.
- `boolean equalsTreePartial(AST t)`: comprueba si t es una versión ampliada del árbol que llama al método.
- *boolean equalsList(AST t)*: igual que `equalsTree` pero admite como entrada una lista de árboles (un árbol sin raíz).

- *boolean equalsListPartial(AST t)*: igual que el método `equalsTreePartial` pero admite como entrada una lista de árboles.
- *String getText()*: obtiene el texto asociado a la raíz.
- *int getType()*: obtiene el tipo asociado a la raíz (habitualmente será un tipo de token).
- *String toString()*: convierte la raíz del árbol a una cadena.
- *String toStringTree()*: convierte el árbol completo en una cadena.
- *String toStringList()*: igual que `toStringTree` pero admite como entrada una lista de árboles.

Con todos estos métodos ya podemos empezar a manipular árboles, lo único que necesitamos es acceder al árbol construido durante el análisis sintáctico, eso lo conseguiremos gracias al método *getAST* de la clase *Anasint*. Podemos ver un ejemplo de uso en el listado A.3. En dicho listado podemos observar primeramente una función principal en la cual se lee el fichero y se prepara para poder ser analizado. Con la función `recorridoArbol` vamos procesando el árbol sintáctico generado de forma recursiva y vamos comprobando en el nodo que estamos si es un nodo `MULTDIV` o un nodo `SUMARESTA` o por el contrario es el caso base que nos encontramos un número, `NUMERO`.

A.4. ANTLRWorks

ANTLRWorks es una aplicación realizada en Java la cual nos facilita el desarrollo de una gramática y su respectivo AST. La aplicación la podemos descargar desde la página oficial de *ANTLR* [13]. Una vez descargada, podemos arrancarla abriendo una terminal y ejecutando el comando `java -jar antlrworks-1.2.1.jar`.

Para arrancar la aplicación, es necesario tener instalado el JDK de Java, nos la podemos descargar en la página oficial de Oracle [17] en la cual también se encuentran las instrucciones de instalación. Una vez arrancada la aplicación, podemos crear nuestra

Listado A.3: Ejemplo de uso del AST

```
1  void calculaArbol(){
2    EjemploLexer lex = new EjemploLexer(new ANTLRFileStream("./Archivo.txt"));
3    CommnTokenStream tokens = new CommonTokenStream(lex);
4    EjemploParser parse = new EjemploParser(tokens);
5    EjemploParser.spec_return spec = parse.spec();
6    CommonTree tree = spec.getTree();
7    final int resultado = recorrdioArbol(tree);
8    system.out.println("El resultado es" + resultado);
9  }
10
11 int recorridoArbol(CommonTree tree){
12   if(tree.getToken().getType() == EjemploParser.NUMERO){
13     final CommonTree variable = (CommonTree) tree.getChild(0);
14     return Integer.parseInt(variable.getText());
15   } else {
16     if (tree.getToken().getType() == EjemploParser.MULTDIV){
17       final CommonTree variable1 = (CommonTree) tree.getChild(0);
18       final CommonTree variable2 = (CommonTree) tree.getChild(1);
19       if(tree.getToken.getText().equals("*")){
20         return recorridoArbol(variable1) * recorridoArbol(variable2);
21       }else{
22         return recorridoArbol(variable1) / recorridoArbol(variable2);
23       }
24     } else if (tree.getToken().getType() == EjemploParser.SUMARESTA) {
25       final CommonTree variable1 = (CommonTree) tree.getChild(0);
26       final CommonTree variable2 = (CommonTree) tree.getChild(1);
27       if(tree.getToken.getText().equals("+")){
28         return recorridoArbol(variable1) + recorridoArbol(variable2);
29       }else{
30         return recorridoArbol(variable1) - recorridoArbol(variable2);
31       }
32     }
33   }
34 }
```

gramática. Para crear la gramática nos dirigimos a `File/New` y creamos el fichero en el cual vamos a escribir la gramática. Podemos ver un ejemplo de cómo quedaría en la aplicación en la figura A.4.

Una vez realizada la gramática podemos depurarla, para ello hacemos clic en `Run/Debug`. Nos aparecerá una pantalla cómo la mostrada en la figura A.5 donde podemos escribir una entrada de prueba.

Una vez compilada, podemos pulsar sobre el botón `AST` para ver el AST que genera dicha entrada y usamos los botones con forma de flecha para ver cómo quedaría dicho AST. En la figura A.6 podemos ver un ejemplo de cómo se muestra en la aplicación.

Las opciones dentro de *ANTLRWorks* son amplias y variadas, en esta sección hemos visto las más utilizada. Como último detalle destacar que *ANTLRWorks* nos puede generar código que implementa el AST: para ello, pulsamos sobre `Generate/Generate Code` y nos creará las clases Java necesarias para la implementación del AST.

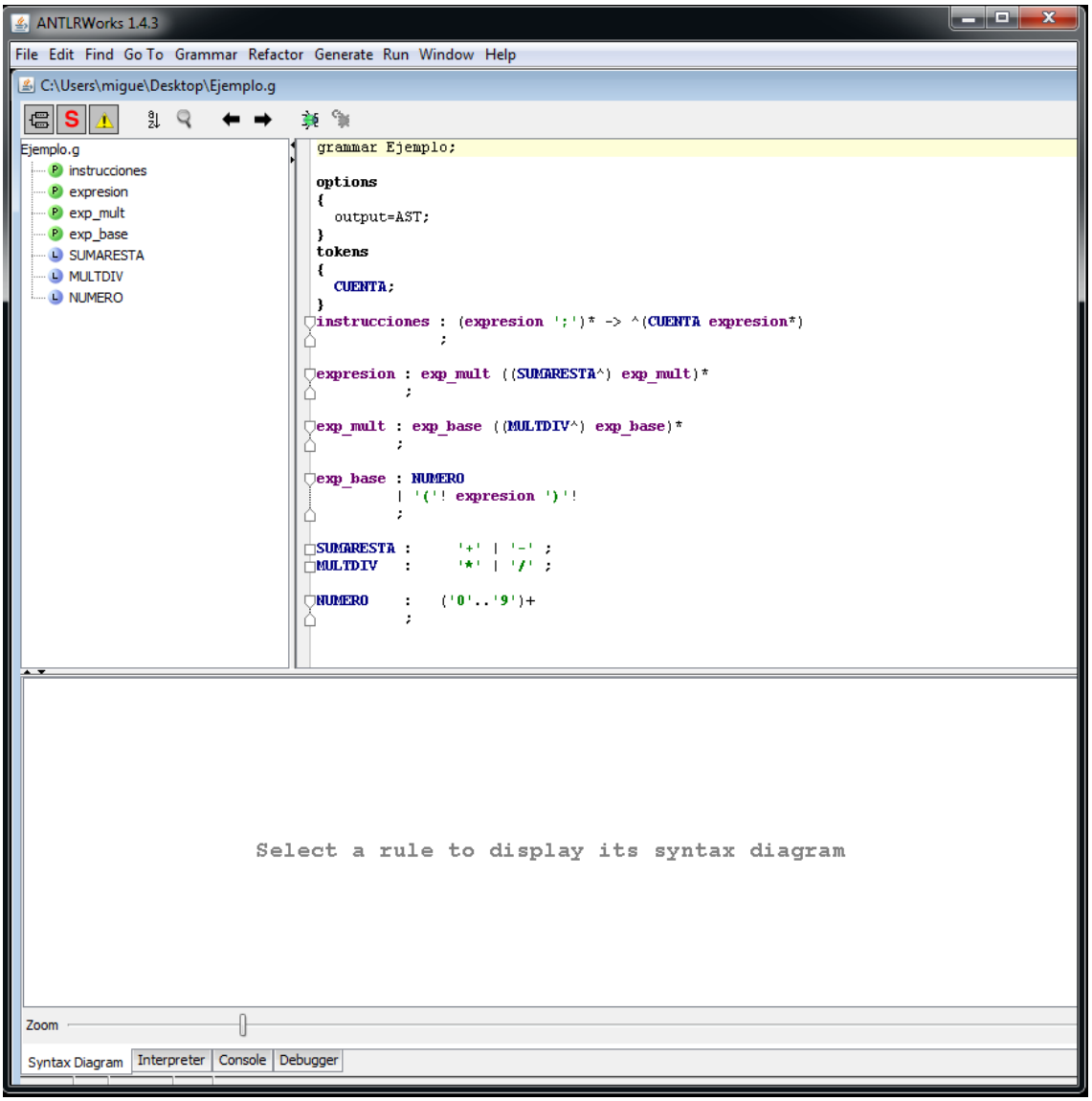

Figura A.4.: Pantalla principal de *ANTLRWorks*

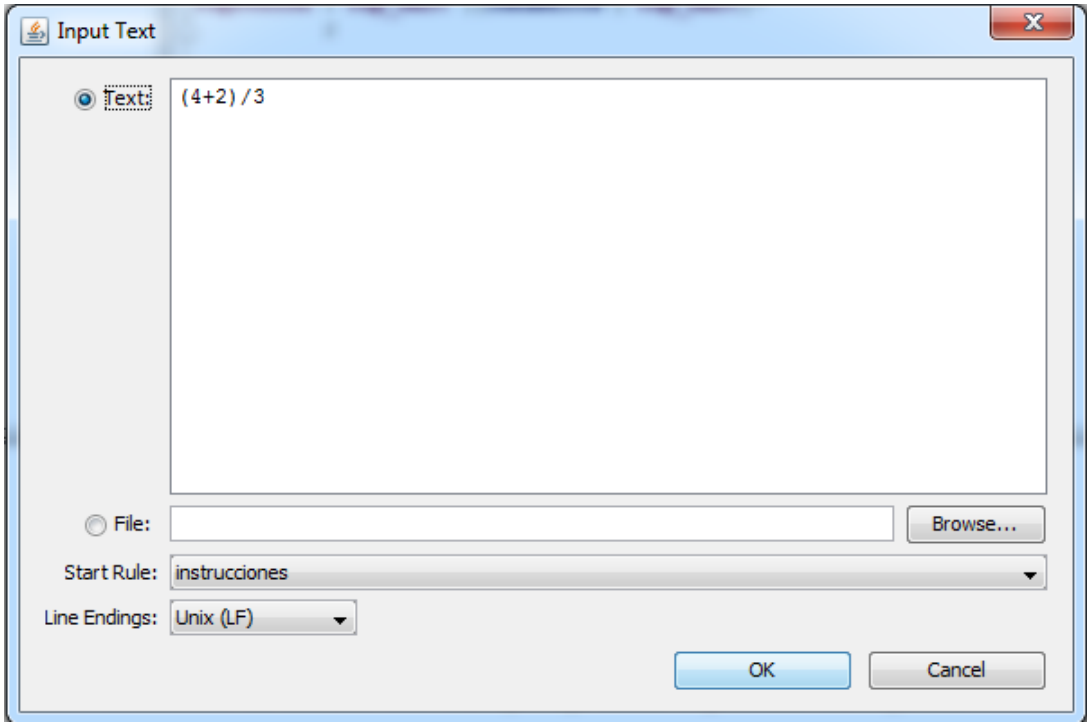

Figura A.5.: Pantalla entrada de un test

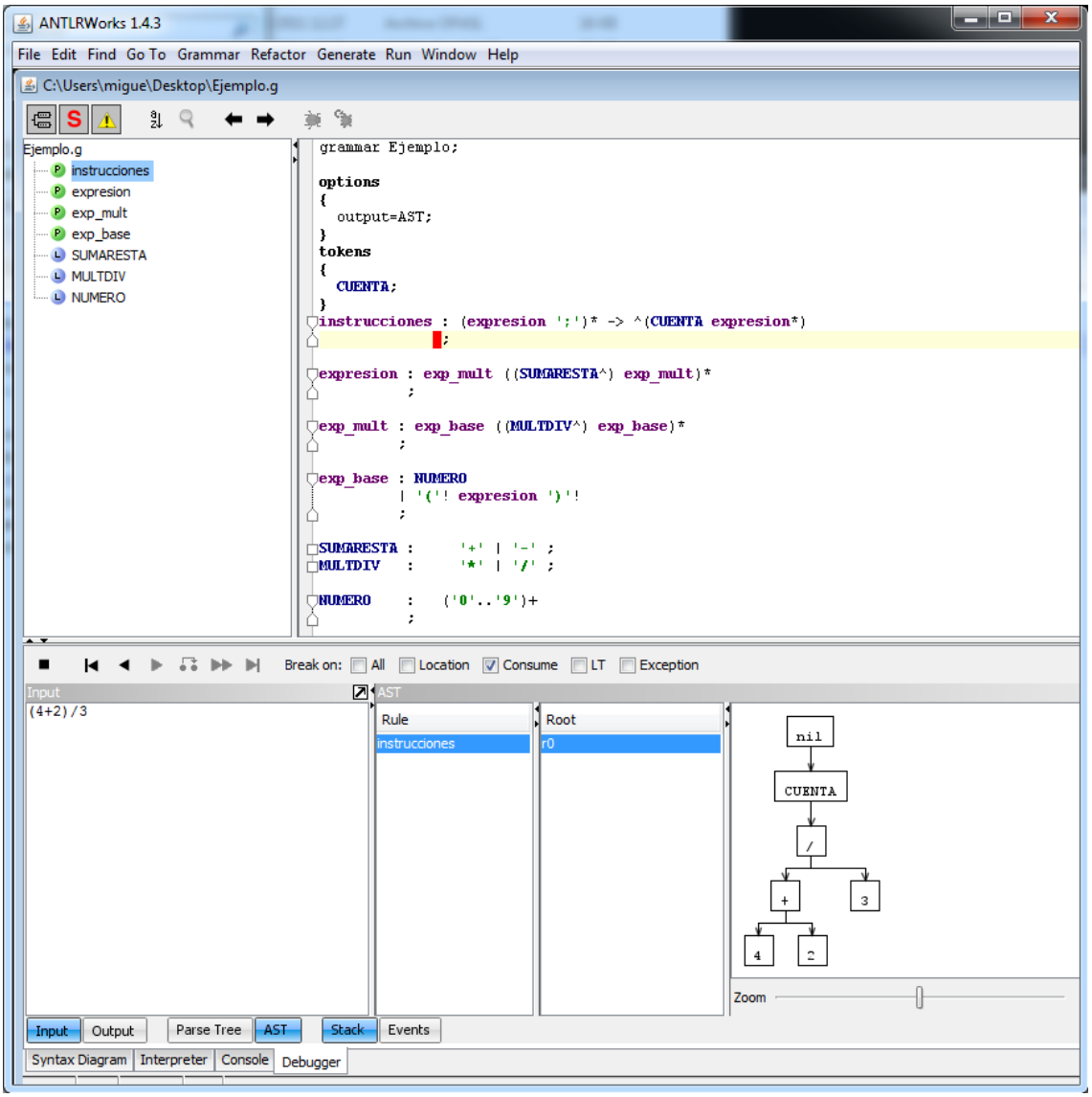

Figura A.6.: Generación visual del AST

Apéndice B

Manual de usuario

En este manual hablaremos de los pasos necesarios para poder usar la herramienta, así como poder escribir adecuadamente especificaciones TestSpec.

La herramienta está destinada a personas con conocimientos sobre pruebas unitarias del software.

B.1. Instalación de TestGenerator

Para instalarse la herramienta bajo una distribución *Linux*, concretamente en *Ubuntu 11.10*, hay que seguir los siguientes pasos:

1. Descargarse el archivo *dist.zip* del enlace `http://bit.ly/y0O2Ej`
2. Descomprimimos el archivo en el directorio que deseemos.
3. Añadimos dicho directorio a la variable `path` del sistema operativo. Si por ejemplo la carpeta la hemos descomprimido en el escritorio, introducimos la siguiente orden en una consola:
   `PATH="$PATH:/home/usuario/Escritorio/test-generator-1.0-SNAPSHOT"`
4. Para que los cambios sean permanente, el usuario debería añadir una línea más al final de *.bashrc*, con este aspecto:

```
export PATH=$PATH: /Escritorio/test-generator-1.0-SNAPSHOT
```

5. Por último tendremos que cerrar sesión y volver a entrar para que los cambios surtan efecto.

Ya está la herramienta lista para usarse.

B.2. Uso de la herramienta

Para poder usar la herramienta, es necesario abrir una consola y escribir la siguiente sentencia:

`testgenerator (argumentos)`

Donde "argumentos" puede tomar una de las siguientes formas:

- `-help`
- `archivo.wsdl servicio operación (in|out|err) [Número de test] [-formateador]`
- `archivo.spec [Número de test] [-formateador]`

Si recibe como argumento `-help`, se mostrará la ayuda de la aplicación.

En el caso que reciba un archivo *wsdl*, tendrá que especificar el servicio del archivo, que operación dentro del servicio y elegir el tipo (`in`, `out`, `err`), de forma opcional puede elegir el número de datos aleatorios que desea que genere, por defecto son cinco, y el formato. Para el formato puede elegir entre `-csv` o `-velocity`, por defecto se usará el formato velocity.

Para el caso que reciba un archivo *spec* sólo tendrá que especificar si lo desea, el número de datos a generar y en que formato desea que muestre el resultado.

B.3. Creación de ficheros TestSpec

TestSpec es el lenguaje especialmente diseñado para realizar especificaciones abstractas del formato de los datos generados por *TestGenerator*.

Una especificación TestSpec tiene dos tipos de sentencias. Mediante el primer tipo se definen los tipos con sus restricciones (sentencias «typedef»). En el segundo, se declaran las variables de cada tipo (sentencias «declaration»).

B.3.1. Sentencias «typedef»

Las sentencias «typedef» vienen estereotipadas de la siguiente forma:

```
typedef Tipo (r1 = valor, r2 = valor...) NombreDefinicion;
```

Dónde:

- typedef: es una palabra reservada de la gramática.
- restricción (r): restricciones que tendrá el tipo anteriormente especificado.
- valor: corresponde con la asignación que tendrá la restricción. Si es única vendrá representado por una sola sentencia, si corresponde con un conjunto de valores aparecerán entre llaves y separadas con comas valor1, valor2...
- NombreDefinicion: identificador asignado al typedef.

Los tipos que podemos definir (sección «typedef») son los siguientes:

- string
- int
- float
- date
- time
- dateTime
- duration
- list

- tuple

A continuación veremos cuáles son las restricciones opcionales u obligatorias que deben tener cada tipo:

- element: esta restricción es obligatoria si el tipo es alguno de los contenedores, representa el tipo de elemento que va a contener. En el caso de las tuplas podríamos necesitar una lista de tipos, por lo que el valor de `element` es una lista ordenada de cadenas separadas por coma.
- min: Este atributo tiene un significado u otro en función del tipo al que se aplique. Aplicado a tipos numéricos, representa el límite inferior inclusivo del espacio de valores del tipo. En el caso de que se aplique a un tipo cadena, indica la longitud mínima que ésta ha de tener. Sin embargo, si es un tipo list, `min` indica el número mínimo de elementos que puede contener la lista.
- max: análogamente a la definición de `min` se le puede aplicar sustituyendo la palabra mínimo por máximo.
- values: Restringe el espacio de valores de un determinado tipo al conjunto de valores especificados.
- pattern: Restringe el espacio de valores de un determinado tipo, restringiendo el espacio léxico a literales que siguen un determinado patrón. El valor debe ser una expresión regular. Esta restricción sólo es aplicable al tipo de dato `string`.
- fractionDigits: controla el número de decimales que deberá contener `float`.
- totalDigits: controla el número total de dígitos que tendrá un número.

Se podrá usar un typedef definido como tipo de un typedef que sea posterior.

B.3.2. Sentencias «declaration»

En las sentencias para realizar declaraciones, aparecerá el nombre de una de las definiciones realizadas en el conjunto de instrucciones «typedef» y el identificador que le

vamos a otorgar a la variable, seguido del carácter fin de línea ';'.

Además, se podrá realizar declaraciones de los tipos básicos int, float y string sin que éstos estén definidos en las sentencias «typedef» anteriormente definidas.

B.4. Ejemplo de un fichero *spec*

Podemos ver un ejemplo en el listado B.1. En el que podemos apreciar en las líneas 1–6 el conjunto de sentencias typedef donde:

- La línea 1 define un tipo cadena denominado boolean que podrá tener como valores las palabras "true" o "false".
- En la línea 2 podemos ver un typedef con identificador *CaraDado* que es un entero que podrá tener los valores en el rango [1,6].
- La línea 3 es una tupla donde el primer elemento es una cadena y el segundo es del tipo *CaraDado* y este typedef se identificará como *ResultadoTirada*.
- En la línea 4 se define un tipo de lista de enteros con 1 elemento como mínimo.
- En la línea 5 se define un tipo entero con valor mínimo de 0, al que identificaremos como `positiveNumber`.
- En la línea 6 se especializa el tipo `positiveNumber` dándole de valor máximo 100 y llamándole `smallNumber`.

En las sentencias de las declaraciones, podemos ver que se declaran cuatro variables de los tipos "boolean", "ResultadoTirada", "ListaNoVaciaInt" y "smallNumber".

Listado B.1: Ejemplo de especificación TestSpec

```
1 typedef string (values={"true", "false"}) boolean;
2 typedef int (min=1, max=6) CaraDado;
3 typedef tuple (element = {string, caraDado}) ResultadoTirada;
4 typedef list (min=1, element = int) ListaNoVaciaInt;
5 typedef int (min=0) positiveNumber;
```

```
6  typedef positiveNumber (max=100) smallNumber;
7
8  boolean myBoolean;
9  ResultadoTirada tirada;
10 ListaNoVaciaInt numeros;
11 smallNumber myNumber;
```

Apéndice C

Manual de desarrollador

En este manual hablaremos de los pasos necesarios que se deberán llevar a cabo si desea descargarse el código fuente y como poder compilarlo.

Este manual está elaborado y probado bajo una distribución *Linux*, concretamente bajo *Ubuntu 11.10*.

C.1. Requisitos del sistema

Para poder obtener el código fuente, debemos tener conexión a Internet en el equipo y tener instalado *Subversion*.

Para instalar *Subversion*, abrimos una terminal y escribimos lo siguiente:

```
sudo apt-get install subversion
```

Los requisitos necesarios para poder compilar el sistema son tener instalado JDK además de *Maven* y *JUnit*, para ellos abrimos una terminal y escribiremos la siguiente orden:

```
sudo apt-get install openjdk-6-jdk maven junit
```

C.2. Obtención del código

Para poder obtener el código fuente del repositorio *SVN*, abrimos una terminal y nos vamos al directorio donde queremos alojar el código. Luego escribimos la siguiente or-

den:

```
svn co https://neptuno.uca.es/svn/sources-fm/trunk/src/test-generator/
```

Con esto obtenemos una copia local de la última versión del código de *TestGenerator*.

Ya es posible navegar por el proyecto, para explorar el código nos dirigiremos a la carpeta *src*. La estructura de cómo está organizado el código es la siguiente:

main/java Directorio principal que contiene el código Java de la aplicación.

test/java Contiene el código de las pruebas.

main/resources Directorio que contiene los recursos utilizados por el código principal.

test/resources Alberga los recursos necesarios para ejecutar las pruebas.

main/assembly Aquí se encuentra los descriptores de distribuciones.

A la hora de desarrollar, le serán útiles los objetivos predefinidos de *Maven*. A continuación se exponen los principales:

clean Elimina todos los directorios de despliegue del proyecto.

compile Compila el código fuente del proyecto.

deploy Despliega el paquete resultante en un repositorio central para ser compartido.

install Instala el paquete en el repositorio local.

package Crea un paquete con el proyecto a partir de las clases compiladas y recursos.

site Genera un sitio con la documentación del proyecto.

test Ejecuta las pruebas o test del proyecto.

Para ejecutar alguno de estos objetivos, abrimos la terminal y nos situamos en el directorio del proyecto. A continuación escribimos la orden en la terminal de la siguiente forma:

```
mvn objetivo
```

Trabajar directamente desde un editor de texto la manipulación de ficheros Java puede ser una tarea bastante engorrosa. Por ello se recomienda usar un IDE como puede ser *NetBeans* o *Eclipse* usando algún «plugin» de integración con *Maven*. Estos «plugins» permiten usar *Maven* desde una interfaz más amigable, evitando así la línea de órdenes y contando con las funcionalidades de refactorización y notificación de errores y avisos de compilación en directo que ofrecen ambos IDE. Las características añadidas a estos entornos son:

- Construir proyectos *Maven* desde el IDE.
- Gestión de dependencias basadas en la sincronización con el pom.xml asociado al proyecto.
- Resolución de dependencias *Maven* en el espacio de trabajo sin necesidad de instalar en repositorios *Maven* locales.
- Descarga automática de las dependencias requeridas desde los repositorios *Maven* remotos.
- Asistentes para crear nuevos proyectos *Maven*, editar los ficheros pom.xml, así como para permitir soporte *Maven* en proyectos ya existentes.
- Búsqueda rápida de dependencias en los repositorios remotos de Maven.
- Avisos en el editor Java para buscar las dependencias o ficheros JAR por el nombre de la clase o del paquete.
- Integración con otras herramientas del IDE de desarrollo elegido.

En el caso de elegir *NetBeans*, desde la versión 6.7 la integración con *Maven* está incorporada en la instalación estándar. En versiones anteriores basta con instalar el «plugin» desde el menú («Tools/Plugins»).

Bibliografía

[1] E. Blanco Muñoz, A. García Domínguez, J. J. Domínguez Jiménez y I. Medina Bulo. Propuesta de una arquitectura para la generación de mutantes de orden superior en WS-BPEL. En *Actas de las XVI Jornadas de Ingeniería del Software y Bases de Datos*, páginas 537–542. A Coruña, España, septiembre 2011.
URL http://www.sistedes.es/jornadas2011/jisbd.htm

[2] J. J. Domínguez Jiménez, A. Estero Botaro, A. García Domínguez y I. Medina-Bulo. GAmera: an automatic mutant generation system for WS-BPEL compositions. En *Proceedings of the 7th IEEE European Conference on Web Services*, páginas 97–106. Eindhoven, Países Bajos, noviembre 2009.

[3] C. Jiménez Gavilán, A. García Domínguez y J. J. Domínguez Jiménez. *Analizador de Servicios Web basados en WSDL 1.1 para pruebas paramétricas*. Proyecto fin de carrera, University of Cádiz, Cádiz, España, mayo 2011.
URL http://hdl.handle.net/10498/11695

[4] R. Chinnici, J. Moreau, A. Ryman y S. Weerawarana. Web Services Description Language (WSDL) version 2.0 part 1: Core language. Informe técnico, W3C, junio 2007. Recomendación W3C.
URL http://www.w3.org/TR/wsdl20

[5] P. V. Biron y A. Malhotra. XML Schema part 2: Datatypes. Recomendación W3C,

World Wide Web Consortium, octubre 2004.
URL http://www.w3.org/TR/xmlschema-2/

[6] K. Ballinger, C. Ferris, M. Gudgin, C. Kevin Liu, M. Nottingham y P. Yendluri. Basic profile - version 1.1 (Final). Informe técnico, Web Services Interoperability Organization, agosto 2004.
URL http://www.ws-i.org/Profiles/BasicProfile-1.1-2004-08-24.html

[7] M. Fowler. Domain-specific languages: An introductory example. http://www.informit.com/articles/article.aspx?p=1592379, septiembre 2010. Fecha de última comprobación: 21 de noviembre de 2011.

[8] Pagina oficial de JCheck. http://www.jcheck.org/. Fecha de última comprobación: 1 de diciembre de 2011.

[9] Pagina oficial de QuickCheck. http://java.net/projects/quickcheck/pages/Home. Fecha de última comprobación: 1 de diciembre de 2011.

[10] B. Eckel. *Piensa en Java*. Pearson Education, 2007.

[11] Pagina oficial de JUnit. http://www.junit.org/.

[12] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad y M. Stal. *Pattern-Oriented Software Architecture: A System Of Patterns*. Wiley, 1996.

[13] J. Bovet. Página oficial de ANTLRWorks. http://www.antlr.org/works/, agosto 2011. Fecha de última comprobación: 24 de octubre de 2011.

[14] E. Gamma, R. Helm, R. Jhonson y J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 2003.

[15] G. Aburruzaga García, I. Medina Bulo y F. Palomo Lozano. *Fundamentos de C++*. Universidad de Cádiz, 2009.

[16] J. A. J. Millán. *Compiladores y procesadores de lenguajes*. Universidad de Cádiz, 2004.

[17] Pagina oficial de Oracle. http://www.oracle.com/technetwork/java/javase/downloads/index.html.

GNU Free Documentation License

Version 1.3, 3 November 2008

<http://fsf.org/>

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to

software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under

this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "**publisher**" means any person or entity that distributes copies of the Document to the public.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in

another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally

prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any,

be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the

"History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes

a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various docu-

ments with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and

disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to

the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-

BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with . . . Texts.” line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.